Estimation, Forecasting, and Policy Analysis with DSGE and State Space Models

Marco Del Negro
Federal Reserve Bank of New York

Banco Central de la República Argentina, Dec 18-19, 2017

IADB Workshop on Macroeconomic Modeling with Julia

Outline

1. Basic notions of Bayesian econometrics
    - Bayes theorem; model selection

2. State-Space models
    - State-space models; the Kalman filter and the likelihood computation; Bayesian estimation; Kalman smoothing and shock decomposition; Carter and Kohn and Durbin and Koopman simulation smoothers, Applications: Time varying parameter models; Factor models; Stochastic Volatility.

3. An introduction to MCMC methods
    - Metropolis-Hastings; Gibbs sampler.

Outline – continued

4. DSGEs
    - Introducing a simple workhorse DSGE model; Estimation; forecasting; impulse response function and variance decomposition.

This is a three hours version of a week long course. The full set of slides for the course is available here.

Main references

Lectures are based on

- Del Negro & Schorfheide, Bayesian Macroeconometrics, Geweke, Koop, and van Dijk (eds.) The Oxford Handbook of Bayesian Econometrics, 2011, Oxford University Press, 293-389. (available on Frank Schorfheide's web page)
- John Geweke, Contemporary Bayesian Econometrics and Statistics, Wiley & Sons, 2005
- Ed Herbst and Frank Schorfheide. Bayesian Estimation of DSGE Models. Princeton University Press. 2015.
- James D. Hamilton, Time Series Analysis, Princeton University Press, 1994.
- An & Schorfheide, Bayesian Analysis of DSGE Models, Econometric Reviews, 26(2-4), 2007, 113-172

Some language

- $y_{1:T} = \{y_1, .., y_t, .., y_T\}$: data (sometimes Y for short), $Y \in \mathcal{Y}$. When not obvious, we will distinguish between the random variable y_t and its realization y_t^o
- θ: parameters, possibly including latent variables; with $\theta \in \Theta$
- $p(y_{1:T}|\theta)$: is the distribution of the data given the parameters (a parametric model); e.g.

$$\mathcal{M}_1 : \quad y_t = \mu + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma^2)$$

  where $\theta = \{\mu, \sigma\}$ and $\Theta = I\!R \times I\!R^+$

$$\Rightarrow \text{pdf of } y_{1:T} \text{ is: } \; p(y_{1:T}|\theta, \mathcal{M}_1) = \Pi_{t=1}^T (2\pi\sigma^2)^{-\frac{1}{2}} \exp(-\frac{1}{2}\frac{(y_t - \mu)^2}{\sigma^2})$$

- Likelihood function: $p(y_{1:T}^o|\theta, \mathcal{M}_1)$ viewed as a function of θ, e.g. $L(\theta; y_{1:T}^o, \mathcal{M}_1) \propto p(y_{1:T}^o|\theta, \mathcal{M}_1)$
- Many models: $\mathcal{M}_i$, e.g. $\mathcal{M}_2 : y_t = \mu + \rho y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma^2)$,...

- Questions:
  - The inference problem: What can I learn about θ from the observed data $y^o_{1:T}$?
  - How can I find out whether the data come from model $\mathcal{M}_1$, or $\mathcal{M}_2$, or ...?

- Bayesian approach: both data Y and parameters θ are random
  - θ is random: the prior $p(\theta)$ reflects my uncertainty about θ before seeing the data
  - the posterior $p(\theta|Y)$ reflects my uncertainty about θ after seeing the data
  - $p(\theta) \rightarrow p(\theta|Y)$? Bayes' law
  - Game plan is simple: form beliefs (probabilities) over what you want to conduct inference on, and update them in light of the data using Bayes law.

Bayes' law

- Given two events A and B, with joint probability $p(A, B)$ and marginals $p(A)$ and $p(B)$:

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

- Similarly:

$$p(\theta|y_{1:T}) = \frac{p(\theta, y_{1:T})}{p(y_{1:T})} = \frac{p(y_{1:T}|\theta)p(\theta)}{p(y_{1:T})}$$

- How do I get $p(y_{1:T})$ (marginal likelihood)?

$$p(y_{1:T}) = \int p(\theta, y_{1:T})d\theta = \int p(y_{1:T}|\theta)p(\theta)d\theta$$

- Conditional on observed data, the posterior distribution and marginal likelihood are $p(\theta|y^o_{1:T})$ and $p(y^o_{1:T})$, respectively.

- Any function $k(\theta|y^o_{1:T}) \propto p(\theta|y^o_{1:T})$ is a (posterior density) kernel

- Call w the vector of interest (e.g., forecasts – in which case $w = y_{T+1,..,T+H}$ – etc.), and assume you have a vector of interests density
$$p(w|y_{1:T}, \theta, \mathcal{M}_i)$$

- Then the object of inference is
$$p(w|y^o_{1:T}, \mathcal{M}_i) = \int p(w|y^o_{1:T}, \theta, \mathcal{M}_i)p(\theta|y^o_{1:T}, \mathcal{M}_i)d\theta$$

  (this if we have only one model on the table – we will discuss later the case where there are many models)

Example I

- Say our model $\mathcal{M}$ is of the form

$$y = \theta + \varepsilon, \ \varepsilon \sim N(0, \sigma_l^2)$$

$$\Rightarrow p(y|\theta) = (2\pi\sigma_l^2)^{-\frac{1}{2}} \exp(-\frac{1}{2}\frac{(y-\theta)^2}{\sigma_l^2})$$

where the prior on θ is given by

$$\theta \sim N(\mu_p, \sigma_p^2)$$

that is

$$p(\theta) = (2\pi\sigma_p^2)^{-\frac{1}{2}} \exp(-\frac{1}{2}\frac{(\theta-\mu_p)^2}{\sigma_p^2})$$

with μ_p, σ_l^2, and σ_p^2 being known quantities.

Hence the joint is

$$
\begin{aligned}
p(y,\theta) &= p(y|\theta)p(\theta) = ((2\pi)^2\sigma_I^2\sigma_p^2)^{-\frac{1}{2}} \\
&\exp\left(-\frac{1}{2}\left[(\sigma_I^{-2}+\sigma_p^{-2})\theta^2 - 2(\sigma_I^{-2}y+\sigma_p^{-2}\mu_p)\theta + \sigma_I^{-2}y^2 + \sigma_p^{-2}\mu_p^2\right]\right) \\
&= p(\theta|y)p(y) = N(\mu_\pi,\sigma_\pi^2)p(y)
\end{aligned}
$$

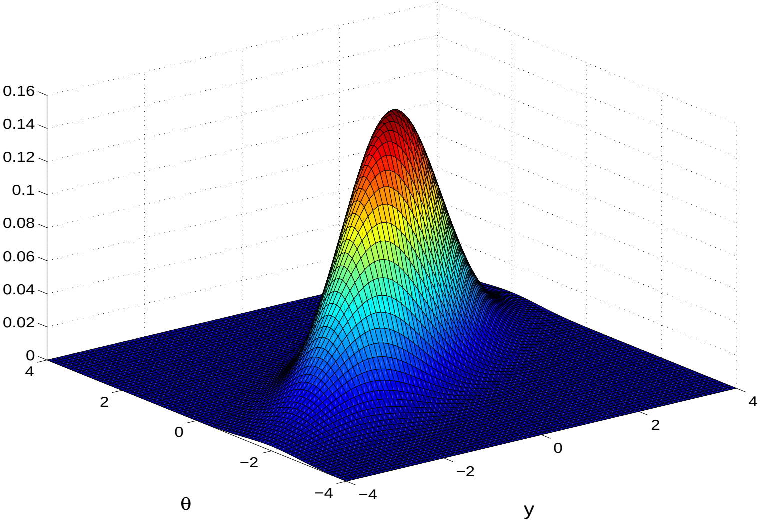
where

$$
\mu_\pi = \frac{\sigma_I^{-2}}{\sigma_I^{-2}+\sigma_p^{-2}}y + \frac{\sigma_p^{-2}}{\sigma_I^{-2}+\sigma_p^{-2}}\mu_p, \quad \sigma_\pi^2 = \left(\sigma_I^{-2}+\sigma_p^{-2}\right)^{-1}
$$

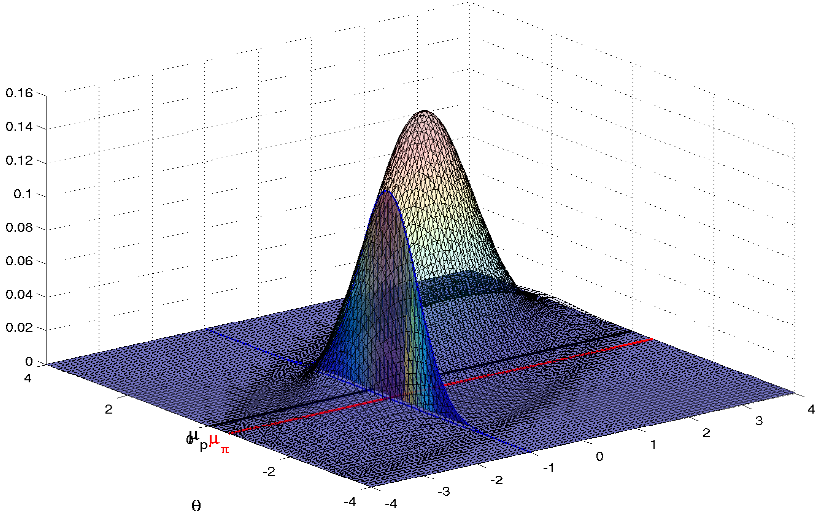
and

$$
\begin{aligned}
p(y) &= (2\pi\frac{\sigma_I^2\sigma_p^2}{\sigma_\pi^2})^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\left[\sigma_I^{-2}y^2+\sigma_p^{-2}\mu_p^2-\sigma_\pi^{-2}\mu_\pi^2\right]\right) \\
&= \left(2\pi\left(\sigma_I^2+\sigma_p^2\right)\right)^{-\frac{1}{2}}\exp(-\frac{1}{2}[\sigma_I^{-2}(1-\frac{\sigma_I^{-2}}{\sigma_I^{-2}+\sigma_p^{-2}})y^2+ \\
&\qquad \sigma_p^{-2}(1-\frac{\sigma_p^{-2}}{\sigma_I^{-2}+\sigma_p^{-2}})\mu_p^2 - 2\frac{\sigma_p^{-2}\sigma_I^{-2}}{\sigma_I^{-2}+\sigma_p^{-2}}y\mu_p])
\end{aligned}
$$

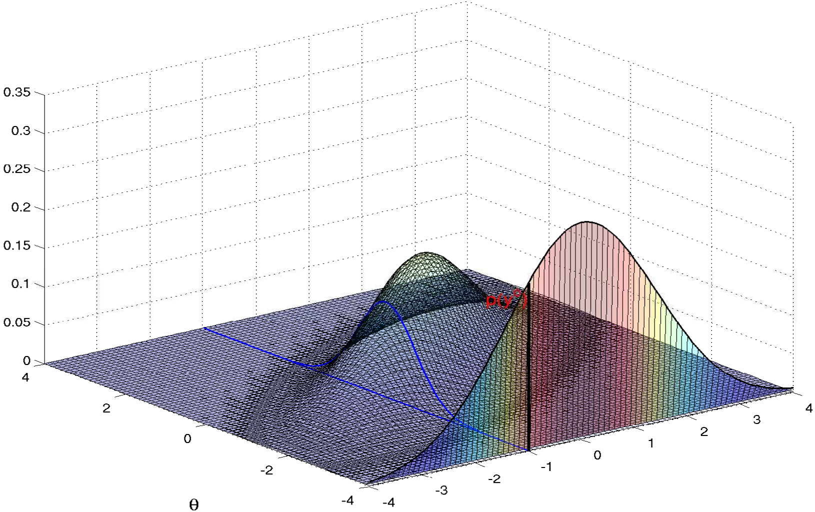
- What's the idea? Prior $\times$ model (likelihood) deliver a joint: $p(\theta, y) = p(y|\theta)p(\theta)$
(this is the $\mu_p = 0$, $\sigma_p = \sigma_l = 1$ case)



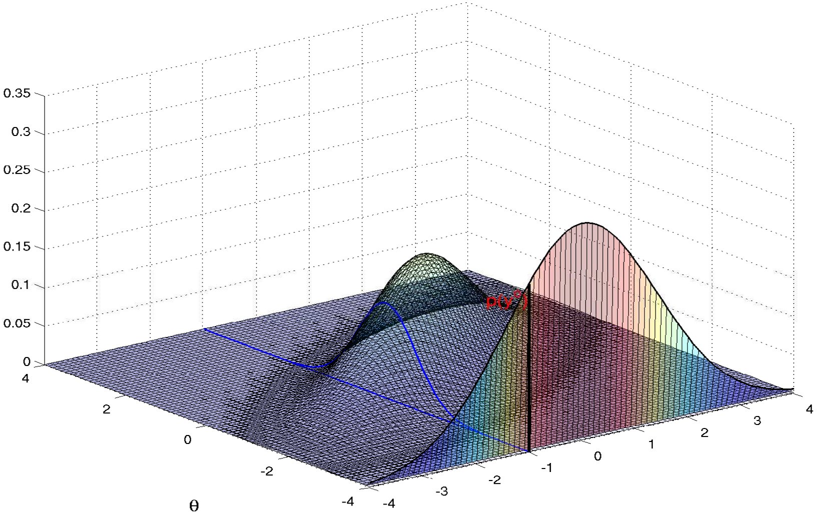
- Now, I observe $y^o_{1:T}$, say $y^o = -1$.
- What's the distribution of θ given that observation? The conditional, which is $\propto$ to the joint computed for $y = y^o_1$: $p(\theta|y^o_1) \propto p(\theta, y^o_1)$



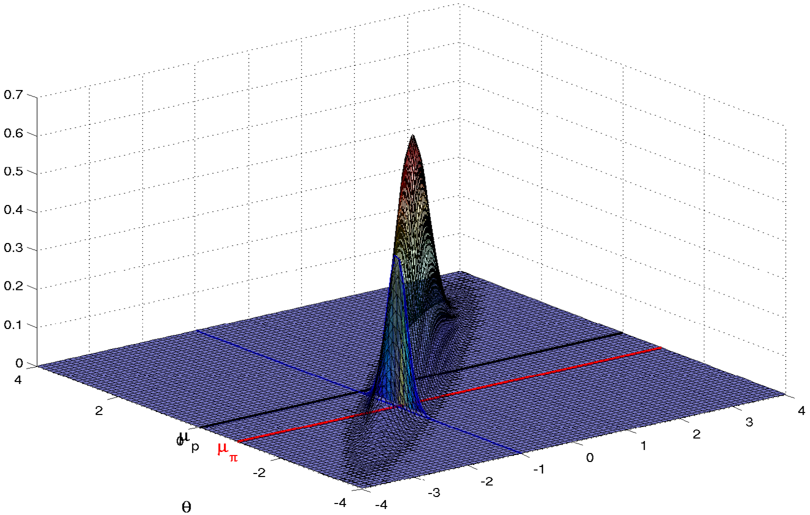
- What is the marginal likelihood? Simply the marginal for y: $p(y_{1:T}) = \int p(\theta, y_{1:T}) d\theta$



- $p(y_1^o)$ is the answer to: How likely was I to observe y_1^0 given the model I had? (and also the normalization constant for the posterior: $p(\theta|y_1^o) = \dfrac{p(\theta, y_1^o)}{p(y_1^o)}$)



- What if the data are more informative ($\sigma_l = \sigma_p/4$)?



Model comparison, Bayes factors, and posterior odds

- Say we are considering two models, $\mathcal{M}_1$ and $\mathcal{M}_2$. Which one fits the data best?

- $p(\mathcal{M}_1)$ and $p(\mathcal{M}_2)$ are our prior probabilities on the two models, with $\sum_{i=1}^{2} p(\mathcal{M}_i) = 1$

- What is the probability of model $\mathcal{M}_i$ after looking at the data?

$$p(\mathcal{M}_i | y^o_{1:T}) = \frac{p(y^o_{1:T} | \mathcal{M}_i) p(\mathcal{M}_i)}{p(y^o_{1:T})}$$

  where $p(y^o_{1:T} | \mathcal{M}_i)$ is the marginal likelihood of model $\mathcal{M}_i$, and $p(y^o_{1:T}) = \sum_{i=1}^{2} p(y^o_{1:T} | \mathcal{M}_i) p(\mathcal{M}_i)$

- So what is the relative probability of model $\mathcal{M}_1$ vs $\mathcal{M}_2$?

$$\underbrace{\frac{p(\mathcal{M}_1|y^o_{1:T})}{p(\mathcal{M}_2|y^o_{1:T})}}_{\text{Posterior Odds}} = \underbrace{\frac{p(y^o_{1:T}|\mathcal{M}_1)}{p(y^o_{1:T}|\mathcal{M}_2)}}_{\text{Bayes Factor}} \underbrace{\frac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)}}_{\text{Prior Odds}}$$

- Why do we care? Because if we have to make decisions about our vector of interest w, which is model-dependent, then we want to figure out how to weight the different models :

$$p(w|y^o_{1:T}) = \sum_i p(w|y^o_{1:T}, \mathcal{M}_i) p(\mathcal{M}_i|y^o_{1:T})$$

- Note: the marginal likelihood is – when normalized – a probability! Posterior odds are ... odds: they capture all remaining uncertainty we have on the relative goodness of fit of model $\mathcal{M}_1$ vs $\mathcal{M}_2$ after observing the data.

- Model $\mathcal{M}_1$:

$$p(y|\theta, \mathcal{M}_1): \quad y = \theta + \varepsilon, \ \varepsilon \sim N(0,1)$$

$$Pr(\theta|\mathcal{M}_1): \quad p(\theta) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{\theta^2}{2})$$

- Model $\mathcal{M}_2$:

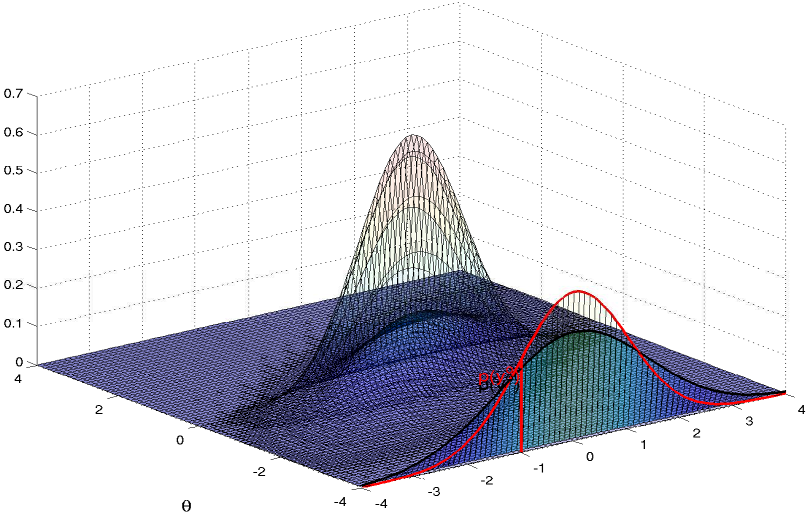
$$p(y|\theta, \mathcal{M}_2): \quad y = \theta + \varepsilon, \ \varepsilon \sim N(0,1)$$

$$Pr(\theta|\mathcal{M}_2): \quad Pr(\theta) = \begin{cases} 1 \text{ if } \theta = 0 \\ 0 \text{ otherwise} \end{cases}$$

- Posterior:

$$\begin{aligned} p(\theta|y, \mathcal{M}_1) &= N(\frac{y}{2}, \frac{1}{2}) \\ Pr(\theta|\mathcal{M}_2) &= \begin{cases} 1 \text{ if } \theta = 0 \\ 0 \text{ otherwise} \end{cases} \end{aligned}$$

- $\mathcal{M}_1$ (black) vs $\mathcal{M}_2$ (red)



- Marginal likelihoods:

$$p(y|\mathcal{M}_1) = (4\pi)^{-\frac{1}{2}} \exp(-\frac{y^2}{4})$$
$$p(y|\mathcal{M}_2) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{y^2}{2})$$

- Bayes factor:

$$\frac{p(y|\mathcal{M}_1)}{p(y|\mathcal{M}_2)} = (2)^{-\frac{1}{2}} \exp(\frac{y^2}{4}) > 1, \text{ if } y^2 > 2\log(2)$$

- Model $\mathcal{M}_3$:

$$p(y|\theta, \mathcal{M}_3): \quad y = \varepsilon, \ \varepsilon \sim N(0,1)$$
$$p(\theta, \mathcal{M}_3): \qquad p(\theta) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{\theta^2}{2})$$

- Same marginal likelihood as $\mathcal{M}_2$: $p(y|\mathcal{M}_3) = p(y|\mathcal{M}_2)$, but very different posterior:

$$p(\theta|y, \mathcal{M}_3) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{\theta^2}{2})$$

Marginal Likelihood as out-of-sample concept

- The marginal data density addresses the question: What is model $\mathcal{M}_i$'s a priori (hence, out-of-sample) guess for what the data are going to look like?

$$p(y_{1:T}|\mathcal{M}_i) \;=\; \int p(y_{1:T}|\,\theta, \mathcal{M}_i)p(\theta|\mathcal{M}_i)d\theta$$

  ... and how does such a guess compare with what the data turned out to be? Marginal likelihood is the likelihood of observing the data under model $\mathcal{M}_i$: $p(y^o_{1:T}|\mathcal{M}_i)$

- .. Or, how well did model $\mathcal{M}_i$ predict the data $y^o_{1:T}$?
- Models that make sharper predictions – if such predictions are not at odds with the data $y^o_{1:T}$ – are favored ($\Rightarrow$ penalty for over-parameterization).

- Predictive density: $p(y_{T+1:T+H}|y^o_{1:T}, \mathcal{M}_i)$
- Predictive likelihood: $p(y^o_{T+1:T+H}|y^o_{1:T}, \mathcal{M}_i)$

- Marginal likelihood is the product of predictive densities – obtained after recursively updating (at every t) the prior/posterior!

$$\begin{aligned} p(y^o_{1:T}|\mathcal{M}_i) &= p(y^o_T|y^o_{1:T-1}, \mathcal{M}_i)p(y^o_{1:T-1}|\mathcal{M}_i) \\ &= \Pi^T_{t=2} p(y^o_t|y^o_{1:t-1}, \mathcal{M}_i)p(y^o_1|\mathcal{M}_i) \end{aligned}$$

  where $p(y^o_1|\mathcal{M}_i) = \int p(y^o_1|\ \theta, \mathcal{M}_i)p(\theta|\mathcal{M}_i)d\theta$ is the predictive likelihood for y^o_1 obtained using the prior.
- M. Friedman ("The Methodology of Positive Economics," 1953) *Theory is to be judged by its predictive power . . . The only relevant test of the validity of a hypothesis is comparison of its predictions with experience.*
- `Exercise: Show that the overall posterior` $p(\theta|y_{1:T})$ `is obtained by recursive updating, that is, at each step` t `you start from the` $t-1$ `posterior` $p(\theta|y_{1:t-1})$ `and update it using the likelihood` $p(y_t|y_{1:t-1}, \theta)$.

State-space models

- Transition equation:

$$s_t = T(\theta)s_{t-1} + R(\theta)\varepsilon_t, \ t = 1, .., T$$

where s_t is $k \times 1$, ε_t is $r \times 1$, θ is a vector of model parameters, and $T(\theta)$ $(k \times k)$ and $R(\theta)$ $(k \times r)$ are functions of these parameters. E.g.

$$s_t = \theta_1 s_{t-1} + \theta_2 \varepsilon_t$$

where simply $T(\theta) = \theta_1$ and $R(\theta) = \theta_2$.

- Measurement equation:

$$y_t = Z(\theta)s_t + D(\theta) + u_t, \ t = 1, .., T$$

where y_t is $n \times 1$, $Z(\theta)$ is $n \times k$ and $D(\theta)$ is $n \times 1$. E.g.

$$y_t = \theta_3 + \theta_4 s_t + u_t$$

where $Z(\theta) = \theta_4$ and $D(\theta) = \theta_3$.

- Distribution of the shocks (ε_t)/measurement error (u_t):

$$\varepsilon_t \sim N(0, Q(\theta)) \; iid, \;\; Q(\theta) \text{ diagonal}; \quad u_t \sim N(0, H(\theta)) \; iid$$

  where $Q(\theta)$ is a diagonal matrix with the σ^2s of each shock on the diagonal (although you do not have to impose this condition on what follows). We will also assume in the derivations that $\boldsymbol{E}[u_s \varepsilon_t'] = 0$, all s, t, although again it is straightforward to derive formulas that allow for correlation.

- Initial conditions:

$$s_0 \sim N(s_{0|0}, P_{0|0})$$

$p(y_{1:T}|\theta)$ for state-space models

- We want to compute

$$p(y_{1:T}|\theta) = p(y_T, ..., y_1|\theta)$$

- Use conditioning!

$$\begin{aligned} p(y_T, ..., y_1|\theta) &= p(y_T|y_{T-1}, ..., y_1, \theta)p(y_{T-1}, ..., y_1, \theta) \\ &= p(y_T|y_{1:T-1}, \theta)..p(y_t|y_{1:t-1}, \theta)..p(y1|\theta) \\ &= \prod_{t=1}^{T} p(y_t|y_{1:t-1}, \theta) \end{aligned}$$

where $y_{1:0} = \{\}$ (i.e., $p(y_1|y_{1:0}, \theta)$ is the unconditional probability).

- But $p(y_t|y_{1:t-1},\theta)$ is Gaussian, and the Gaussian distribution is fully nailed down by its mean and variance.

- If we know $y_{t|t-1} = E(y_t|y_{1:t-1},\theta)$ and $V_{t|t-1} = Var(y_t|y_{1:t-1},\theta)$ we can compute

$$\begin{aligned} p(y_t|y_{1:t-1},\theta) &= (2\pi)^{-\frac{1}{2}}|V_{t|t-1}|^{-\frac{1}{2}} \\ &\qquad \exp\left(-\frac{1}{2}(y_t - y_{t|t-1})' V_{t|t-1}^{-1}(y_t - y_{t|t-1}\right) \end{aligned}$$

and hence $p(y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|y_{1:t-1},\theta)$

How do we get $y_{t|t-1}$ and $V_{t|t-1}$? Kalman filter!

- The Kalman filter is a recursive algorithm.
- Say you know
  $s_{t-1/t-1} = E(s_{t-1}|y_{1:t-1}, \theta), \; P_{t-1/t-1} = Var(s_{t-1}|y_{1:t-1}, \theta)$

$$\begin{matrix} s_{t-1/t-1} \\ P_{t-1/t-1} \end{matrix} \xrightarrow{\text{forecasting}} \begin{matrix} s_{t/t-1} \\ P_{t/t-1} \end{matrix} \rightarrow \begin{matrix} y_{t/t-1} \\ V_{t/t-1} \end{matrix} \xrightarrow{\text{update}} \begin{matrix} s_{t/t} \\ P_{t/t} \end{matrix}$$

- Forecasting:
  1. Use

$$s_t = T(\theta)s_{t-1} + R(\theta)\varepsilon_t$$

  to obtain

$$s_{t|t-1} = Ts_{t-1|t-1}$$

$$P_{t|t-1} = TP_{t-1|t-1}T' + RQR'$$

  2. Use

$$y_t = Z(\theta)s_t + D(\theta) + u_t, \; t = 1, .., T$$

  to obtain

$$y_{t|t-1} = Zs_{t|t-1} + D$$

$$V_{t|t-1} = ZP_{t|t-1}Z' + H$$

- Updating
- An aside on conditional distribution for Gaussian variables (*normal updating*). Say y and s are jointly Gaussian

$$\begin{matrix} y \\ s \end{matrix} \sim N \left(\begin{matrix} \mu_y \\ \mu_s \end{matrix} \begin{bmatrix} \Sigma_{yy} & \Sigma_{ys} \\ \Sigma'_{ys} & \Sigma_{ss} \end{bmatrix} \right)$$

... then here's how you get the conditional distribution

$$E[s|y] = \mu_s + \Sigma'_{ys}\Sigma_{yy}^{-1}(y - \mu_y)$$

$$V[s|y] = \Sigma_{ss} - \Sigma'_{ys}\Sigma_{yy}^{-1}\Sigma_{ys}$$

- By the same token, since the distribution of s_t and y_t conditional on $t-1$ information is

$$\begin{matrix} y_t \\ s_t \end{matrix} \Bigg| y_{1:t-1} \sim N \left(\begin{matrix} y_{t|t-1} \\ s_{t|t-1} \end{matrix} \begin{bmatrix} V_{t|t-1} & ZP_{t|t-1} \\ P_{t|t-1}Z' & P_{t|t-1} \end{bmatrix} \right)$$

$$s_{t|t} = s_{t|t-1} + P'_{t|t-1}Z'V_{t|t-1}^{-1}(y_t - y_{t|t-1})$$

$$P_{t|t} = P_{t|t-1} - P'_{t|t-1}Z'V_{t|t-1}^{-1}ZP_{t|t-1}$$

- How do we start the algorithm? Recall we assumed

$$s_0 \sim N(s_{0|0}, P_{0|0})$$

- How do we choose $s_{0|0}, P_{0|0}$? If s_t is stationary, a natural choice is the ergodic distribution: $s_{0|0} = E[s_t] = 0$, and $P_{0|0} = E[s_t s_t']$ solves the Lyapunov equation

$$P_{0|0} = TP_{0|0}T' + RQR'$$

- Note that

$$s_{t+1|t} = Ts_{t|t-1} + K_t(y_t - y_{t|t-1})$$
$$P_{t+1|t} = TP_{t|t-1}T' - TP'_{t|t-1}Z'K'_t + RQR'$$

where $K_t = TP'_{t|t-1}Z'V^{-1}_{t|t-1}$ is called the Kalman gain.

- Recursive formulation for $P_{t+1|t}$

$$\begin{aligned}
P_{t+1|t} &= TP_{t|t-1}T' - TP'_{t|t-1}Z'K'_t + RQR' \\
&= TP_{t|t-1}T' - TP'_{t|t-1}Z'V^{-1}_{t|t-1}ZP_{t|t-1}T' + RQR' \\
&= TP'_{t|t-1}(I - Z'(ZP_{t|t-1}Z' + H)^{-1}ZP_{t|t-1})T' + RQR'
\end{aligned}$$

for $t \to \infty$, $P_{t+1|t} \to \bar{P}_{1|0}$ and $K_t \to \bar{K}$

$$\begin{aligned}
s_{t+1|t} &= Ts_{t|t-1} + \bar{K}(y_t - y_{t|t-1}) \\
&= \bar{K}(y_t - D) - \bar{K}(y_{t|t-1} - D) + Ts_{t|t-1} \\
&= \bar{K}(y_t - D) + (T - \bar{K}Z)s_{t|t-1} \\
&= \sum_{j=0}^{\infty}(T - \bar{K}Z)^j\bar{K}(y_{t-j} - D)
\end{aligned}$$

Innovation representation

- $t|t-1 \;\to t+1|t$
- Define the innovations

$$x_t = s_t - s_{t/t-1}$$

and the forecast errors

$$\nu_t = y_t - y_{t/t-1} = Zx_t + u_t$$

- Define $L_t = T - K_t Z$, then

$$\begin{aligned} x_{t+1} &= Ts_t + R\varepsilon_{t+1} - Ts_{t|t-1} - K_t\nu_t \\ &= Tx_t - K_t Z x_t + R\varepsilon_{t+1} \\ &= L_t x_t + R\varepsilon_{t+1} \end{aligned}$$

- These formulas are the "innovation analogue" of the state-space model. An alternative updating formula for $P_{t+1|t}$ is:

$$P_{t+1|t} = TP_{t|t-1}L_t' + RQR',$$

and the whole Kalman filter recursion can be defined in terms of ν_t, $s_{t|t-1}$, $P_{t|t-1}$, the formula for $V_{t|t-1}$ and the matrices K_t and L_t.

- Used in Koopman, Disturbance smoother for state space models, Biometrika 1993. Here are some notes of mine (and Jenny Chan, Dan Greenwald) explaining Koopman's smoother using our notation, and some Matlab code implementing it (see `kalsmth_k93.m`).

Learning about latent variables

- Address questions like
    - What are the drivers of business cycles? What shocks caused the Great Recession? ...
    - How large is the output gap $(\hat{y}_t - \hat{y}_t^f)$?
- Want to draw from

$$p(s_{0:T}|\theta, y_{1:T})$$

- ... but the shocks are not part of $s_{1:t}$! Just add them: Create the variables s_t^ε, defined by

$$s_t^\varepsilon = \varepsilon_t,$$

and stack them to s_t: $\tilde{s}_t = [s_t, s_t^\varepsilon]$. The new tansition equation is:

$$\tilde{s}_t = \tilde{T}(\theta)\tilde{s}_{t-1} + \tilde{R}(\theta)\varepsilon_t$$

where $\tilde{T}(\theta)$ and $\tilde{R}(\theta)$ are adjusted to accommodate s_t^ε .

- In terms of Bayesian updating the joint distribution of data and unobservables (parameters and latent variables) is given by:

$$p(y_{1:T}, s_{0:T}, \theta) = \underbrace{p(y_{1:T}|s_{0:T}, \theta)}_{\text{measurement}} \underbrace{p(s_{0:T}|\theta)}_{\text{transition}} p(\theta)$$

- We integrate out the states $s_{0:T}$ (Kalman filter):

$$p(y_{1:T}|\theta)p(\theta) = \left(\underbrace{\int p(y_{1:T}|s_{0:T}, \theta)p(s_{0:T}|\theta)ds_{0:T}}_{p(y_{1:T}|\theta)} \right) p(\theta)$$

- ... and write the joint posterior of $\theta, s_{0:T} \mid y_{1:T}$ as marginal times conditional:

$$p(s_{0:T}, \theta|y_{1:T}) = p(s_{0:T}|\theta, y_{1:T})p(\theta|y_{1:T})$$

Smoothing and simulation smoothers

- How do we draw from $p(s_{0:T}|\theta, y_{1:T})$? Realize that (omitting θ from the conditioning to simplify notation)

$$\begin{aligned}
p(s_{0:T}|y_{1:T}) &= p(s_0|s_{1:T}, y_{1,T})p(s_{1:T}|y_{1,T}) \\
&= \left[\prod_{t=0}^{T-1} p(s_t|s_{t+1:T}, y_{1:T})\right] p(s_T|y_{1:T}) \\
&= \left[\prod_{t=0}^{T-1} p(s_t|s_{t+1:T}, y_{1:t})\right] p(s_T|y_{1:T}) \qquad (*) \\
&= \left[\prod_{t=0}^{T-1} p(s_t|s_{t+1}, y_{1:t})\right] p(s_T|y_{1:T}) \qquad (**)
\end{aligned}$$

- Step $(*)$: Why is $p(s_t|s_{t+1:T}, y_{1:T}) = p(s_t|s_{t+1:T}, y_{1:t})$? Note that

$$y_{t+j} = Zs_{t+j} + u_{t+j}, \quad j \geq 1.$$

Since $u_{t+j}, \quad j \geq 1$ is independent from $\varepsilon_{t-s}, \quad s \geq 0$, there is no additional information in y_{t+j} about s_t if I already know s_{t+1}.

- Step (**): Why is $p(s_t|s_{t+1:T}, y_{1:t}) = p(s_t|s_{t+1}, y_{1:t})$? Note that

$$s_{t+1+j} = T^j s_{t+1} + \sum_{k=1}^{j} T^{j-k} R\varepsilon_{t+1+j},\ j \geq 1.$$

Call

$$s_{t|s_{t+1},y_{1:t}} = E\left[s_t|s_{t+1}, y_{1:t}\right],\ s_{t+1+j|s_{t+1},y_{1:t}} = E\left[s_{t+1+j}|s_{t+1}, y_{1:t}\right]$$

and realize that *conditional* on $y_{1:t}$, s_t and s_{t+1+j} are uncorrelated and therefore independent (gaussianity):

$$E\left[\left(s_t - s_{t|s_{t+1},y_{1:t}}\right)\left(s_{t+1+j} - s_{t+1+j|s_{t+1},y_{1:t}}\right)'|s_{t+1}, y_{1:t}\right] =$$

$$E\left[\left(s_t - s_{t|s_{t+1},y_{1:t}}\right)\left(\sum_{k=1}^{j} T^{j-k} R\varepsilon_{t+1+j}\right)'|s_{t+1}, y_{1:t}\right] = 0$$

because $E[\varepsilon_{t+1+j}|\varepsilon_{1:t+1}] = 0$ (i.i.d. assumption).

Simulation smoother (Carter and Kohn)

- We have established that

$$p(s_{0:T}|y_{1:T}) = \left[\prod_{t=0}^{T-1} p(s_t|s_{t+1}, y_{1:t})\right] p(s_T|y_{1:T})$$

- This implies that the sequence $s_{1:T}$, conditional on $y_{1:T}$, can be drawn recursively:

  1. Draw s_T from $p(s_T|y_{1:T})$
  2. For $t = T-1, .., 0$, draw s_t from $p(s_t|s_{t+1}, y_{1:t})$

- How do I draw from $p(s_T|y_{1:T})$?
- i) I know that $s_T|y_{1:T}$ is gaussian, ii) I have $s_{T|T} = E[s_T|y_{1:T}]$ and $P_{T|T} = \text{Var}[s_T|y_{1:T}]$ from the filtering procedure $\Rightarrow$

$$s_T|y_{1:T} \sim N\left(s_{T|T}, P_{T|T}\right)$$

- How do we draw from $p(s_t|s_{t+1}, y_{1:t})$? We know that

$$\begin{array}{c} s_{t+1} \\ s_t \end{array} \Bigg| y_{1:t} \sim N\left(\begin{array}{c} s_{t+1|t} \\ s_{t|t} \end{array} \left[\begin{array}{cc} P_{t+1|t} & TP_{t|t} \\ P_{t|t}T' & P_{t|t} \end{array} \right] \right)$$

  Note: 1) easy to show that $E\left[(s_{t+1} - s_{t+1|t})(s_t - s_{t|t})'\right] = TP_{t|t}$, 2) we know all these matrices from the Kalman filter.

- Then ...

$$E\left[s_t|s_{t+1}, y_{1:t}\right] = s_{t|t} + P'_{t|t}T'P^{-1}_{t+1|t}(s_{t+1} - s_{t+1|t})$$

$$\mathrm{Var}\left[s_t|s_{t+1}, y_{1:t}\right] = P_{t|t} - P'_{t|t}T'P^{-1}_{t+1|t}TP_{t|t}$$

- ... and

$$s_t|s_{t+1}, y_{1:t} \sim N\left(E\left[s_t|s_{t+1}, y_{1:t}\right], \mathrm{Var}\left[s_t|s_{t+1}, y_{1:t}\right]\right)$$

Kalman smoothing

- What if I just want to know $s_{t|T} = E[s_t | y_{1:T}]$ and $P_{t|T} = \text{Var}[s_t | y_{1:T}]$? (note $s_{t|T} \neq E[s_t | s_{t+1}, y_{1:t}]$!)

  Two approaches:

  1. If I've run my simulation smoother, I have the draws from the *joint* $p(s_{0:T} | y_{1:T})$: $s^j_{0:T}, \ j = 1, .., n^{sim}$. Take the draws from the *marginal* (namely $s^j_t, \ j = 1, .., n^{sim}$) and compute mean and variance!

  2. Kalman smoothing (from the "old days", when simulation smoothing was computationally challenging). Again, the algorithm is recursive:

     1. Derive a mapping $(s_{t+1|T}, \ P_{t+1|T}) \ \rightarrow \ (s_{t|T}, \ P_{t|T})$

     2. Start from $(s_{T|T}, \ P_{T|T})$ and proceed backwards for $t = T - 1, .., 0$

- Let's derive the mapping:

$$\begin{aligned} s_{t|T} &= E\left[s_t|y_{1:T}\right] \\ &= E\left[E\left[s_t|s_{t+1},y_{1:T}\right]|y_{1:T}\right] && (*) \\ &= E\left[E\left[s_t|s_{t+1},y_{1:t}\right]|y_{1:T}\right] && (**) \\ &= E\left[s_{t|t} + P'_{t|t}T'P^{-1}_{t+1|t}(s_{t+1}-s_{t+1|t})|y_{1:T}\right] && (***) \\ &= s_{t|t} + P'_{t|t}T'P^{-1}_{t+1|t}(s_{t+1|T}-s_{t+1|t}) && (****) \end{aligned}$$

- Step $(*)$: Law of iterated expectations
- Step $(**)$: Given s_{t+1}, $y_{t+1:T}$ contains no additional information about s_t (see discussion above)
- Step $(***)$: Plug in formula obtained before
- Step $(****)$: All $.|t$ variables are *known* given $y_{1:T}$ (since information spanned by $y_{1:t}$ is contained in the information set spanned by $y_{1:T}$)

- Similarly

$$
\begin{aligned}
P_{t|T} &= E\left[(s_t - s_{t|T})(s_t - s_{t|T})'|y_{1:T}\right] \\
&= E\left[s_t s_t'|y_{1:T}\right] - s_{t|T}s_{t|T}' && (*) \\
&= E\left[E\left[s_t s_t'|s_{t+1}, y_{1:T}\right]|y_{1:T}\right] - s_{t|T}s_{t|T}' && (**) \\
&= E\left[E\left[s_t s_t'|s_{t+1}, y_{1:t}\right]|y_{1:T}\right] - s_{t|T}s_{t|T}' && (***) \\
&= P_{t|t} - P_{t|t}' T' P_{t+1|t}^{-1}\left(P_{t+1|t} - P_{t+1|T}\right) P_{t+1|t}^{-1} T P_{t|t} && (****)
\end{aligned}
$$

- Step $(*)$: $\text{Var}(x) = E(x^2) - E(x)^2$
- Step $(**)$ and $(***)$: same as before
- Homework: you figure out $(****)$
  hint: realize that $s_{t+1} - s_{t+1|t} = s_{t+1} - s_{t+1|T} + s_{t+1|T} - s_{t+1|t}$

Some references

Kalman filter/smoother:

- relevant chapter in James D. Hamilton. 1994. Time Series Analysis. Princeton University Press

Books on simulation smoothers/state-space models:

- James Durbin and Siem Jan Koopman. 2001. Time Series Analysis by State Space Methods. Oxford University Press
- Chang-Jin Kim and Charles R. Nelson. 1998. State-Space Models with Regime-Switching: Classical and Gibbs-Sampling Approaches with Applications. MIT Press
- Giordani, P., M.K. Pitt, and R. Kohn (2011), "Bayesian Inference for Time Series State Space Models." In J. Geweke, G. Koop, and H. van Dijk (eds.), Handbook of Bayesian Econometrics, Oxford University Press

Fast smoothers: The idea

Durbin and Koopman, A simple and efficient for state space time series analysis, Biometrika 2002

- Say you have two normally distributed random variables, x and y. You know how to (i) draw from the joint $p(x, y)$ and (ii) to compute $\boldsymbol{E}[x|y]$.
- You want to generate a draw from $x|y^0 \sim \mathcal{N}(\boldsymbol{E}[x|y^0], W)$ for some y^0. Proceed as follows:

1. Generate a draw (x^+, y^+) from $p(x, y)$.

   By definition, x^+ is also a draw from $p(x|y^+) = \mathcal{N}(\boldsymbol{E}[x|y^+], W)$ or, alternatively, $x^+ - \boldsymbol{E}[x|y^+]$ is a draw from $\mathcal{N}(0, W)$.
2. Use $\boldsymbol{E}[x|y^0] + x^+ - \boldsymbol{E}[x|y^+]$ is a draw from $\mathcal{N}(\boldsymbol{E}[x|y^0], W)$

   Since the variables are normally distributed the scale W *does not depend on the location* y (draw a two dimensional normal, or review the formulas for normal updating, to convince yourself that is the case). Hence $p(x|y^+)$ and $p(x|y^0)$ have the same variance W, which means that $\boldsymbol{E}[x|y^0] + x^+ - \boldsymbol{E}[x|y^+]$ is a draw from $\mathcal{N}(\boldsymbol{E}[x|y^0], W)$.

Fast smoothers

- Imagine you know how to compute the smoothed estimates of the shocks $\boldsymbol{E}[\varepsilon_{1:T}|y_{1:T}]$ (see Koopman, Disturbance smoother for state space models, Biometrika 1993)
- ... and want to obtain draws from $p(\varepsilon_{1:T}|y_{1:T})$ (again, we omit θ for notational simplicity). Proceed as follows:

1. Generate a new draw $(\varepsilon^+_{1:T}, s^+_{1:T}, y^+_{1:T})$ from $p(\varepsilon_{1:T}, s_{1:T}, y_{1:T})$ by drawing $s_{0|0}$ and $\varepsilon_{1:T}$ from their respective distributions, and then using the transition and measurement equations.
2. Compute $\boldsymbol{E}[\varepsilon_{1:T}|y_{1:T}]$ and $\boldsymbol{E}[\varepsilon_{1:T}|y^+_{1:T}]$ (and $\boldsymbol{E}[s_{1:T}|y_{1:T}]$ and $\boldsymbol{E}[s_{1:T}|y^+_{1:T}]$ if need the states);
3. Compute $\boldsymbol{E}[\varepsilon_{1:T}|y_{1:T}] + \varepsilon^+_{1:T} - \boldsymbol{E}[\varepsilon_{1:T}|y^+_{1:T}]$ (and $\boldsymbol{E}[s_{1:T}|y_{1:T}] + s^+_{1:T} - \boldsymbol{E}[s_{1:T}|y^+_{1:T}]$).

- Refinement: Given that the conditional expectations $\boldsymbol{E}[\varepsilon_{1:T}|y_{1:T}]$ and $\boldsymbol{E}[\varepsilon_{1:T}|y^+_{1:T}]$ are linear in y, steps 1 and 3 can be sped up by computing $\boldsymbol{E}[\varepsilon_{1:T}|y_{1:T} - y^+_{1:T}]$ and then obtaining the draw from $\varepsilon^+_{1:T} + \boldsymbol{E}[\varepsilon_{1:T}|y_{1:T} - y^+_{1:T}]$. The last two steps in the algorithm change as follows:

1. Compute $\boldsymbol{E}[\varepsilon_{1:T}|y^*_{1:T}]$ (and $\boldsymbol{E}[s_{1:T}|y^*_{1:T}]$ if need the states);
2. Compute $\boldsymbol{E}[\varepsilon_{1:T}|y^*_{1:T}] + \varepsilon^+_{1:T}$ (and $\boldsymbol{E}[s_{1:T}|y^*_{1:T}] + s^+_{1:T}$).

- Here is some Matlab code implementing the Durbin Koopman smoother.

Forecasting

- How do we generate forecasts $y_{T+1:T+H}$ from a state-space model? Simple...

$$p(y_{T+1:T+H}|y_{1:T}) =$$
$$\int_{(s_T,\theta)} p(y_{T+1:T+H}|s_T,\theta,y_{1:T}) \underbrace{p(s_T|\theta,y_{1:T})}_{\text{posterior of } s_T|\theta} \underbrace{p(\theta|y_{1:T})}_{\text{posterior of } \theta} d(s_T,\theta)$$

where

$$p(y_{T+1:T+H}|s_T,\theta,y_{1:T}) = \int_{s_{T+1:T+H}} p(y_{T+1:T+H}|s_{T+1:T+H}) \\ p(s_{T+1:T+H}|s_T,\theta,y_{1:T})ds_{T+1:T+H}$$

In words...:

1. Use the Kalman filter to compute mean and variance of the distribution $p(s_T|\theta^{(j)}, y_{1:T})$. Generate a draw $s_T^{(j)}$ from this distribution, where $\theta^{(j)}$ is a draw from the posterior of θ.
2. Draw from $s_{T+1:T+H}|(s_T, \theta, y_{1:T})$ by generating a sequence of innovations $\epsilon^{(j)}_{T+1:T+H}$, and iterating the state transition equation forward starting from $s_T^{(j)}$:

$$s_t^{(j)} = T(\theta^{(j)})s_{t-1}^{(j)} + R(\theta^{(j)})\epsilon_t^{(j)}, \quad t = T+1, \dots, T+H.$$

3. Use the measurement equation to obtain $y^{(j)}_{T+1:T+H}$:

$$y_t^{(j)} = D(\theta^{(j)}) + Z(\theta^{(j)})s_t^{(j)}, \quad t = T+1, \dots, T+H. \quad \square$$

Point forecasts

- Given a loss function $L(y_{T+h}, \hat{y}_{T+h})$, find the prediction that minimizes the posterior expected loss:

$$\hat{y}_{T+h|T} = \text{argmin}_\delta \int_{y_{T+h}} L(y_{T+h}, \delta) p(y_{T+h}|y_{1:T}) dy_{T+h}.$$

- If you have a quadratic loss function

$$L(y_{T+h}, \delta) = tr[(y_{T+h} - \delta)' W (y_{T+h} - \delta)]$$

where W is a symmetric positive-definite weight matrix the optimal predictor is the posterior mean

$$\hat{y}_{T+h|T} = \int_{y_{T+h}} y_{T+h} p(y_{T+h}|y_{1:T}) dy_{T+h} \approx \frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} y_{T+h}^{(j)},$$

Estimation of State-space Models

- Come up with a prior $p(\theta)$.
- Obtain posterior
$$p(\theta|y_{1:T}) \propto p(y_{1:T}|\theta)p(\theta)$$
where $\propto$ comes from the fact that $p(y_{1:T})$ does not depend on θ.
- How do I draw from $p(\theta|y_{1:T})$ when it is unrecognizable? MCMC (Markov Chain Montecarlo) methods!

Simulation Methods

- Say you have a posterior

$$\pi(\theta|y_{1:T}) = p(y_{1:T}|\theta)p(\theta)/p(y_{1:T})$$

  that is is not of known form.

- How do I draw from $\pi(\theta|y_{1:T})$? MCMC (Markov Chain Montecarlo) methods!
  - Monte Carlo methods are a class of computational algorithms that rely on repeated random sampling to compute their results: Use the computer to generate a (very long) sequence of draws $\{\theta^{(1)}, ..., \theta^{(j-1)}, \theta^{(j)}, ..., \theta^{(J)}\}$
  - Markov Chain because the way draws are generated follows a Markov structure: $\theta^{(j-1)} \to \theta^{(j)}$.

Some references

Textbooks:

- Andrew Gelman, John B. Carlin, Hal S. Stern, Donald B. Rubin. Bayesian Data Analysis, Second Edition. Chapman & Hall/CRC Texts in Statistical Science. *comment: great manual for MCMC methods*
- John Geweke. Contemporary Bayesian Econometrics and Statistics. John Wiley & Sons, Inc. 2005. *comment: great overview of Bayesian methods in econometrics, also, discussion of why MCMC works*
- Fabio Canova. Methods for Applied Macroeconomic Research. Princeton University Press. 2007. *comment: overview of quantitative methods in macroeconomics*
- Tony Lancaster. An introduction to modern Bayesian econometrics. Wiley-Blackwell. 2004 *comment: Introduction to Bayesian econometrics*

- Ed Herbst and Frank Schorfheide. Bayesian Estimation of DSGE Models. Princeton University Press. 2015. *comment: Most updated book on DSGE Estimation; chapters on SMC and particle filter*

Articles:

- Chib and Greenberg. Understanding the Metropolis Hastings Algorithm. American Statistician, 49(4), 327335, 1995.
- Chib, "Introduction to Simulation and MCMC Methods," In J. Geweke, G. Koop, and H. van Dijk (eds.), Handbook of Bayesian Econometrics, Oxford University Press

Classical Simulation Methods: Accept-reject

- This is MC (Monte Carlo) but not MC (Markov Chain)
- The goal is to obtain draws from $\pi(\theta)$. Draw θ from a so-called proposal or source density $q(\theta)$ (we drop the conditioning on $y_{1:T}$ for simplicity) which is such that, for all $\theta \in \Theta$:

$$\pi(\theta) \leq cq(\theta)$$

(that is $c = \sup_{\theta \in \Theta} \dfrac{\pi(\theta)}{q(\theta)}$)

- Algorithm: For each iteration $j = 1, .., J$

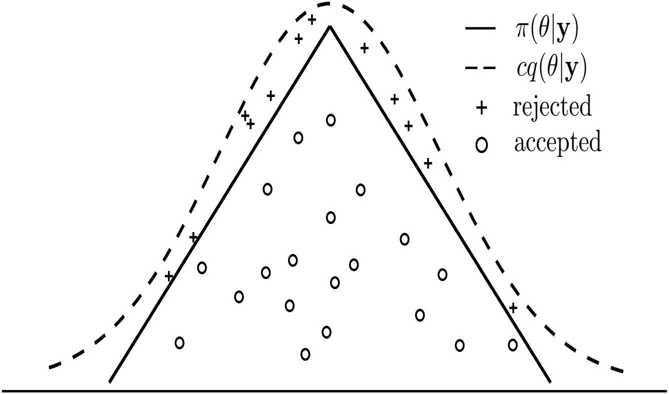
1. Propose $\theta^* \sim q(\theta)$ and $U \sim Unif[0, 1]$
2. Accept-Reject: set $\theta^j = \theta^*$ if

$$U \leq \frac{\pi(\theta)}{cq(\theta)}$$

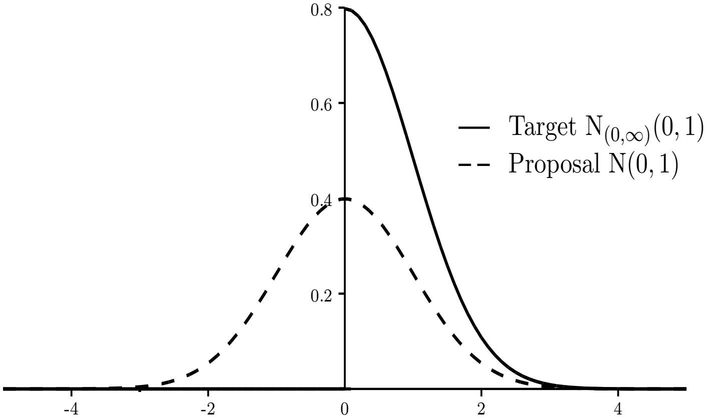
otherwise repeat (1).

3. Collect $\{\theta^{(1)}, ..., \theta^{(J)}\}$

- Dots are $Ucq(\theta)$. Reject if $Ucq(\theta) > \pi(\theta)$.
- Intuition: Reject if the "gap" between proposal $q(\theta)$ and true distribution $\pi(\theta)$ is large.



- Example: drawing from truncated standard normal using standard normal as proposal (note: $c = 2$).



Importance Sampling

- Say you want to compute

$$E_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta$$

where $\pi(\theta) = \dfrac{k(\theta)}{\int k(\theta)d\theta}$ – that is, $k(\theta)$ is the *kernel*.

- That is equal to:

$$E_\pi[h(\theta)] = \frac{\int h(\theta)\frac{k(\theta)}{q(\theta)}q(\theta)d\theta}{\int \frac{k(\theta)}{q(\theta)}q(\theta)d\theta} = \frac{\boldsymbol{E}_q[h(\theta)\frac{k(\theta)}{q(\theta)}]}{\boldsymbol{E}_q[\frac{k(\theta)}{q(\theta)}]}$$

so the following should be a reasonable estimator

$$\bar{h}_J = \frac{1}{J}\sum_j w(\theta^{(j)})h(\theta^{(j)})$$

- Importance sampling works as long as the importance weights

$$w(\theta^{(j)}) = \frac{k(\theta^{(j)})/q(\theta^{(j)})}{\frac{1}{J}\sum_j k(\theta^{(j)})/q(\theta^{(j)})}$$

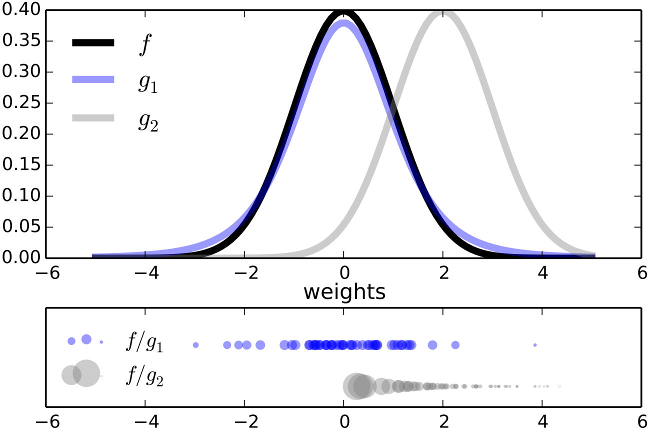
are bounded as a function of θ (see Geweke 2005 for more details).

- Key advantage of importance sampling relative to accept-reject is that in the former you do not have to know/compute the bound (you just have to know that they are bounded), in the latter you have to know c.

Illustration

If θ^i's are draws from $q(\cdot)$ then

$$\mathbb{E}_\pi[h] \approx \frac{\frac{1}{J}\sum_{i=1}^J h(\theta^i)w(\theta^i)}{\frac{1}{J}\sum_{i=1}^J w(\theta^i)}, \quad w(\theta) = \frac{k(\theta)}{q(\theta)}.$$



Accuracy

- Since we are generating *iid* draws from $q(\theta)$, it's fairly straightforward to derive a CLT:

- It can be shown that

$$\sqrt{J}(\bar{h}_J - \mathbb{E}_\pi[h]) \Longrightarrow J\big(0, \Omega(h)\big), \quad \text{where} \quad \Omega(h) = \mathbb{V}_q[(\pi/q)(h - \mathbb{E}_\pi[h])].$$

- Using a crude approximation (see, e.g., Liu (2008)), we can factorize $\Omega(h)$ as follows:

$$\Omega(h) \approx \mathbb{V}_\pi[h]\big(\mathbb{V}_q[\pi/q] + 1\big).$$

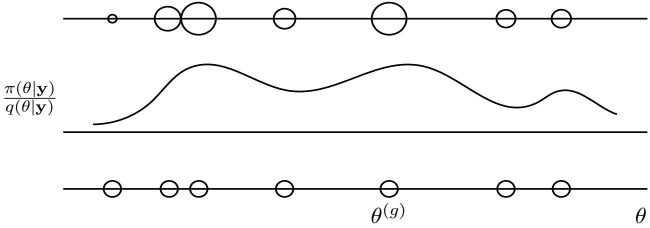
  The approximation highlights that the larger the variance of the importance weights, the less accurate the Monte Carlo approximation relative to the accuracy that could be achieved with an *iid* sample from the posterior.

- Users often monitor

$$ESS = J\frac{\mathbb{V}_\pi[h]}{\Omega(h)} \approx \frac{J}{1 + \mathbb{V}_q[\pi/q]}.$$

Sampling Importance Re-sampling (SIR)

- How can we get *draws* (as opposed to just moments) from $\pi(\theta)$?
- Since $\pi(\theta) = \dfrac{\pi(\theta)}{q(\theta)} q(\theta)$, then if $\{\theta^{(1)}, ..., \theta^{(J)}\}$ are draws from $q(\theta)$ the target can be expressed as the discreet distribution
$$\hat{\pi}(\theta) = w(\theta^{(j)})\delta(\theta - \theta^{(j)})$$
with $\delta(\theta - \theta^{(j)}) = 1$ if $\theta = \theta^{(j)}$ and zero otherwise (Dirac).
- Call $\{\theta^{(1)}, ..., \theta^{(J)}\}$ particles.
- So to get new particles $\{\theta^{*(1)}, ..., \theta^{*(L)}\}$ just resample $\{\theta^{(1)}, ..., \theta^{(J)}\}$ with replacement with probabilities $\{w(\theta^{(j)})\}$.



MCMC

- Imagine you have a (proper, ie. $\int K(\theta, \theta^\dagger) d\theta^\dagger = 1$) transition kernel $K(\theta, \theta^\dagger)$ that is reversible, i.e., that satisfies

$$K(\theta, \theta^\dagger)\pi(\theta) = K(\theta^\dagger, \theta)\pi(\theta^\dagger)$$

(the likelihood of moving from θ to $\theta^\dagger$ is the same as the likelihood of the reverse move)

- ... then it is also invariant, i.e.

$$\pi(\theta^\dagger) = \int K(\theta, \theta^\dagger)\pi(\theta) d\theta$$

(once you have converged to $\pi(\theta)$, you remain in $\pi(\theta)$).

- History. From the question "What does $K(\theta, \theta^\dagger)$ converge to?" to "How can I build a $K(\theta, \theta^\dagger)$converging to $\pi(\theta)$? "

Metropolis-Hastings Algorithm

- Draw θ^* from a so-called proposal density $q(\theta^*|\theta^{(j-1)})$.

- Set $\theta^{(j)} = \theta^*$ with probability

$$\alpha(\theta^*|\theta^{(j-1)}) = \min\left\{1,\ \frac{\pi(\theta^*|y_{1:T})/q(\theta^*|\theta^{(j-1)})}{\pi(\theta^{(j-1)}|y_{1:T})/q(\theta^{(j-1)}|\theta^*)}\right\}$$

  and $\theta^{(j)} = \theta^{(j-1)}$ otherwise.

- Why does MH work – that is, why is it reversible?
- Imagine the case where

$$q(\theta^*|\theta)\pi(\theta) > q(\theta|\theta^*)\pi(\theta^*)$$

(more likely to move $\theta \longrightarrow \theta^*$ than $\theta^* \longrightarrow \theta$)

- We can "correct the flow" by introducing probabilities $\alpha(\theta^*|\theta)$ and $\alpha(\theta|\theta^*)$ such that

$$\alpha(\theta^*|\theta)q(\theta^*|\theta)\pi(\theta) = q(\theta|\theta^*)\pi(\theta^*)\alpha(\theta|\theta^*)$$

- Specifically, make $\alpha(\theta|\theta^*)$ as high as possible ($\alpha(\theta|\theta^*) = 1$) and then choose

$$\alpha(\theta^*|\theta) = \frac{q(\theta|\theta^*)\pi(\theta^*)}{q(\theta^*|\theta)\pi(\theta)}$$

- But since you do not always make a move, the kernel $K_{MH}(\theta, \theta^\dagger)$ has actually two components:

$$K_{MH}(\theta, \theta^\dagger) = \alpha(\theta^\dagger|\theta)q(\theta^\dagger|\theta) + \delta(\theta^\dagger - \theta)r(\theta)$$

where $\delta(\theta^\dagger)$ is the Dirac function

$$\delta(\theta^\dagger - \theta) = \begin{cases} 1 & \text{if } \theta = \theta^\dagger \\ 0 & \text{otherwise.} \end{cases}$$

and

$$r(\theta) = \int (1 - \alpha(\theta^\dagger|\theta))q(\theta^\dagger|\theta)d\theta^\dagger = 1 - \int \alpha(\theta^\dagger|\theta)q(\theta^\dagger|\theta)d\theta^\dagger$$

(note: $\int \alpha(\theta^\dagger|\theta)q(\theta^\dagger|\theta)d\theta^\dagger$ is the average acceptance probability)

- Easy to show that

$$K_{MH}(\theta, \theta^\dagger)\pi(\theta) = K_{MH}(\theta^\dagger, \theta)\pi(\theta^\dagger)$$

since

$$\delta(\theta^\dagger - \theta)r(\theta)\pi(\theta) = \delta(\theta - \theta^\dagger)r(\theta^\dagger)\pi(\theta^\dagger)$$

(both sides $\neq 0$ only when $\theta = \theta^\dagger$)

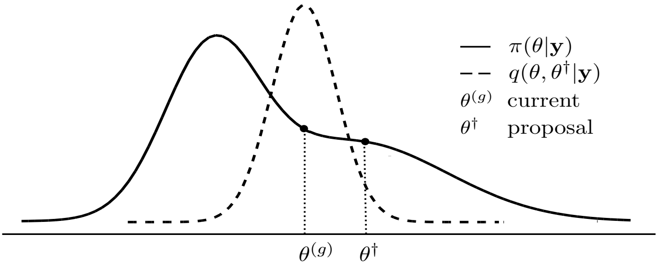
Random Walk Metropolis-Hastings Algorithm

- In Random-Walk Metropolis: $q(\theta^*|\theta^{(j-1)}) = q(\theta^{(j-1)}|\theta^*)$ e.g.

$$\theta^* = \theta^{(j-1)} + N(0, \tilde{V})$$

and the expression simplifies to

$$\alpha(\theta^*|\theta^{(j-1)}) = \min\left\{1,\ \frac{\pi(\theta^*|y_{1:T})}{\pi(\theta^{(j-1)}|y_{1:T})}\right\}$$



Multiple Blocks Metropolis-Hastings Algorithm

- Partition θ into two blocks $\{\theta_1, \theta_2\}$, and devise proposal densities $q(\theta_1^*|\theta_1, \theta_2)$ and $q(\theta_2^*|\theta_1, \theta_2)$. Algorithm:

1. Draw θ_1^* from $q(\theta_1^*|\theta_1, \theta_2)$. Set $\theta_1^{(j)} = \theta_1^*$ with probability

$$\alpha(\theta_1^*|\theta_1^{(j-1)}, \theta_2^{(j-1)}) = \min\left\{1,\ \frac{\pi(\theta_1^*, \theta_2^{(j-1)})/q(\theta_1^*|\theta_1^{(j-1)}, \theta_2^{(j-1)})}{\pi(\theta_1^{(j-1)}, \theta_2^{(j-1)})/q(\theta_1^{(j-1)}|\theta_1^*, \theta_2^{(j-1)})}\right\}$$

   and $\theta_1^{(j)} = \theta_1^{(j-1)}$ otherwise.

2. Draw θ_2^* from $q(\theta_2^*|\theta_1, \theta_2)$. Set $\theta_2^{(j)} = \theta_2^*$ with probability

$$\alpha(\theta_2^*|\theta_1^{(j)}, \theta_2^{(j-1)}) = \min\left\{1,\ \frac{\pi(\theta_1^{(j)}, \theta_2^*)/q(\theta_2^*|\theta_1^{(j)}, \theta_2^{(j-1)})}{\pi(\theta_1^{(j)}, \theta_2^{(j-1)})/q(\theta_2^{(j-1)}|\theta_1^{(j)}, \theta_2^*)}\right\}$$

   and $\theta_2^{(j)} = \theta_2^{(j-1)}$ otherwise.

Gibbs Sampler

- Requirements: Suppose the parameter vector θ can be partitioned into $\theta = [\theta_1', \dots, \theta_m']'$. For each i it is possible to generate draws of θ_i from the conditional distribution $\pi(\theta_i|\theta_{-i}, Y)$ where θ_{-i} denotes the vector θ without the partition θ_i.

- For $j = 1, \dots, J$:

1. Draw $\theta_1^{(j)}$ from the density $\pi(\theta_1|\theta_2^{(j-1)}, \dots, \theta_m^{(s)}, Y)$.
2. Draw $\theta_2^{(j)}$ from the density $\pi(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_m^{(j-1)}, Y)$.
3. $\cdots$
4. Draw $\theta_m^{(j)}$ from the density $\pi(\theta_m|\theta_1^{(j)}, \dots, \theta_{m-1}^{(j)}, Y)$. $\square$

- Why does it work? Think of Gibbs Sampler as a Multiple Move MH with proposals (in the 2 blocks case) $q(\theta_1^*|\theta_1,\theta_2) = \pi(\theta_1|\theta_2)$ and $q(\theta_2^*|\theta_1,\theta_2) = \pi(\theta_2|\theta_1)$.
- Note that
$$\frac{\pi(\theta_1^*,\theta_2^{(j-1)})}{\pi(\theta_1^{(j-1)},\theta_2^{(j-1)})} = \frac{\pi(\theta_1^*|\theta_2^{(j-1)})}{\pi(\theta_1^{(j-1)}|\theta_2^{(j-1)})}$$
and hence
$$\begin{aligned}\alpha(\theta_1^*|\theta_1^{(j-1)},\theta_2^{(j-1)}) &= \min\left\{1, \frac{\pi(\theta_1^*,\theta_2^{(j-1)})/q(\theta_1^*|\theta_1^{(j-1)},\theta_2^{(j-1)})}{\pi(\theta_1^{(j-1)},\theta_2^{(j-1)})/q(\theta_1^{(j-1)}|\theta_1^*,\theta_2^{(j-1)})}\right\}\\ &= \min\left\{1, \frac{\pi(\theta_1^*|\theta_2^{(j-1)})/\pi(\theta_1^*|\theta_2^{(j-1)})}{\pi(\theta_1^{(j-1)}|\theta_2^{(j-1)})/\pi(\theta_1^{(j-1)}|\theta_2^{(j-1)})}\right\}\\ &= 1\end{aligned}$$
- You always accept! Same for the other block.

Another Take on the Gibbs Sampler

- What's the idea? Suppose you want to draw from

$$\pi(\theta_1, \theta_2)$$

  and you don't know how ...

- But you know how to draw from

$$\pi(\theta_1|\theta_2) \propto \pi(\theta_1, \theta_2) \text{ and } \pi(\theta_2|\theta_1) \propto \pi(\theta_1, \theta_2)$$

- Gibbs sampler: you obtain draws from $\pi(\theta_1, \theta_2)$ by drawing repeatedly from $\pi(\theta_1|\theta_2)$ and $\pi(\theta_2|\theta_1)$

Why does it work?

- Some theory of Markov chains.
- Say you want to draw from the marginal $\pi(\theta_1)$ (note, by Bayes' law if you have draws from the marginal you also have draws from the joint $\pi(\theta_1, \theta_2)$).
- If you find a **Markov transition kernel** $K(\theta_1, \theta_1^\dagger)$ that solves the *fixed point integral equation*:

$$\pi(\theta_1^\dagger) = \int K(\theta_1, \theta_1^\dagger)\pi(\theta_1)d\theta_1$$

  (and that is π^*-*irreducible* and *aperiodic*) ...

- Then if you generate draws $\theta_1^{(j)}$, $j = 1, ..., J$ starting from $\theta_1^{(0)}$,

$$|K(A, \theta_1^{(0)})^m - \pi(A)| \to 0 \text{ for any set } A \text{ and any } \theta_1$$

  and

$$\frac{1}{J}\sum_j h(\theta_1^{(j)}) \to \int h(\theta_1)\pi(\theta_1)d\theta_1$$

Why does it work?

- But wait... the Gibbs sample does provide a Markov transition kernel

$$K(\theta_1, \theta_1^\dagger) = \int \pi(\theta_1^\dagger|\theta_2)\pi(\theta_2|\theta_1)d\theta_2$$

- ... that solves the *fixed point integral equation*:

$$\begin{aligned}
\pi(\theta_1^\dagger) &= \int K(\theta_1, \theta_1^\dagger)\pi(\theta_1)d\theta_1 \\
&= \int \left(\int \pi(\theta_1^\dagger|\theta_2)\pi(\theta_2|\theta_1)d\theta_2 \right) \pi(\theta_1)d\theta_1 \\
&= \int \pi(\theta_1^\dagger|\theta_2) \left(\int \pi(\theta_2|\theta_1)\pi(\theta_1)d\theta_1 \right) d\theta_2 \\
&= \int \pi(\theta_1^\dagger|\theta_2)\pi(\theta_2)d\theta_2 = \pi(\theta_1^\dagger)
\end{aligned}$$

(and sufficient conditions for π^*-*irreducibility* and *aperiodicity* are usually met, see Chib and Greenberg 1996).

SMC (Sequential Monte Carlo)

- “Standard” MCMC can be inaccurate, especially in medium and large-scale DSGE models
  - Modify MCMC algorithms to overcome weaknesses: blocking of parameters; tailoring of (mixture) proposal densities

- Sequential Monte Carlo (SMC): (more precisely, sequential importance sampling):
  - Better suited to handle irregular and multimodal posteriors associated with large DSGE models.
  - Algorithms can be easily parallelized.

- SMC = “Importance Sampling on steroids”
  - Theoretical work: Chopin (2004); Del Moral, Doucet, Jasra (2006)
  - Applied work: Creal (2007); Durham and Geweke (2011, 2012)
  - For DSGE applications: Ed Herbst and Frank Schorfheide. Bayesian Estimation of DSGE Models. Princeton University Press. 2015.

From Importance Sampling to Sequential Importance Sampling

- In general, it's hard to construct a good proposal density $q(\theta)$,
- especially if the posterior has several peaks and valleys.
- Idea - Part 1: it might be easier to find a proposal density for

$$\pi_n(\theta) = \frac{[p(Y|\theta)]^{\phi_n} p(\theta)}{\int [p(Y|\theta)]^{\phi_n} p(\theta) d\theta} = \frac{k_n(\theta)}{Z_n}.$$

  at least if ϕ_n is close to zero.
- Idea - Part 2: We can try to turn a proposal density for π_n into a proposal density for π_{n+1} and iterate, letting $\phi_n \longrightarrow \phi_N = 1$.

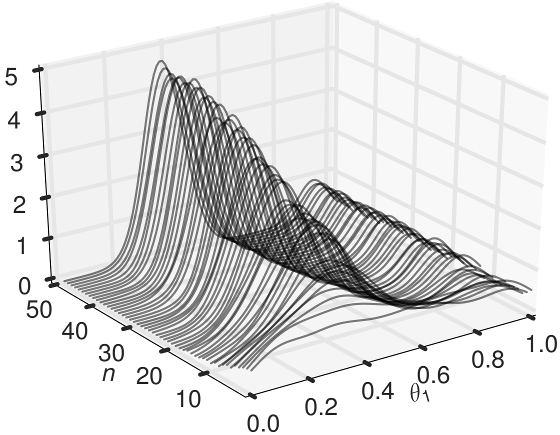
Illustration:

- Our state-space model:

$$y_t = [1\ 1]s_t, \quad s_t = \left[\begin{array}{cc} \theta_1^2 & 0 \\ (1-\theta_1^2) - \theta_1\theta_2 & (1-\theta_1^2) \end{array} \right] s_{t-1} + \left[\begin{array}{c} 1 \\ 0 \end{array} \right] \epsilon_t.$$

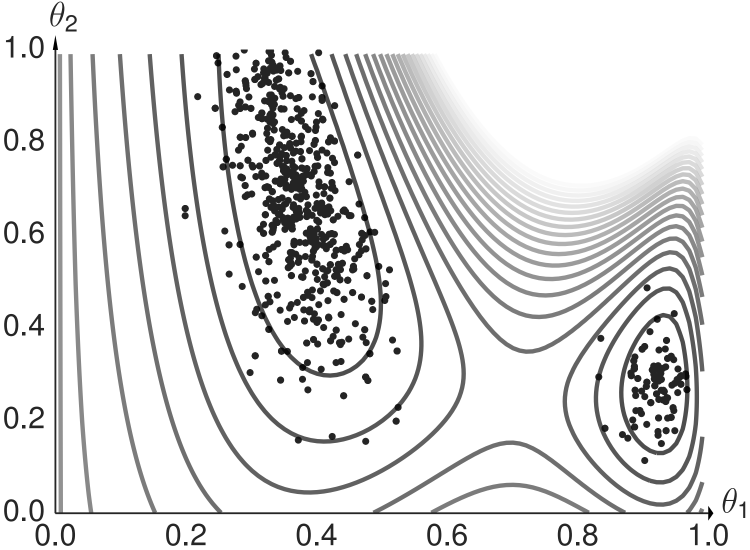
- Innovation: $\epsilon_t \sim iidN(0,1)$.
- Prior: uniform on the square $0 \leq \theta_1 \leq 1$ and $0 \leq \theta_2 \leq 1$.
- Simulate $T = 200$ observations given $\theta = [0.45, 0.45]'$, which is observationally equivalent to $\theta = [0.89, 0.22]'$

Illustration: Tempered Posteriors of θ_1

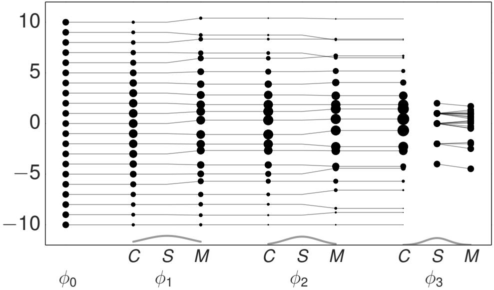


$$\pi_n(\theta) = \frac{[p(Y|\theta)]^{\phi_n} p(\theta)}{\int [p(Y|\theta)]^{\phi_n} p(\theta) d\theta} = \frac{k_n(\theta)}{Z_n}, \quad \phi_n = \left(\frac{n}{N_\phi}\right)^\lambda$$

Illustration: Posterior Draws



SMC Algorithm: A Graphical Illustration



- $\pi_n(\theta)$ is represented by a swarm of particles $\{\theta_n^j, W_n^j\}_{j=1}^J$:

$$\bar{h}_{n,J} = \frac{1}{J}\sum_{j=1}^{J} W_n^j h(\theta_n^j) \xrightarrow{a.s.} \mathbb{E}_{\pi_n}[h(\theta_n)].$$

- C is Correction; S is Selection; and M is Mutation.

SMC Algorithm

1. **Initialization.** ($\phi_0 = 0$). Draw the initial particles from the prior: $\theta_1^i \stackrel{iid}{\sim} p(\theta)$ and $W_1^j = 1$, $j = 1, \ldots, J$.
2. **Recursion.** For $n = 1, \ldots, N_\phi$,
   1. **Correction.** Reweight the particles from stage $n-1$ by defining the incremental weights

      $$\tilde{w}_n^j = [p(Y|\theta_{n-1}^j)]^{\phi_n - \phi_{n-1}}$$

      and the normalized weights

      $$\tilde{W}_n^j = \frac{\tilde{w}_n^j W_{n-1}^j}{\frac{1}{J}\sum_{j=1}^J \tilde{w}_n^j W_{n-1}^j}, \quad j = 1, \ldots, J.$$

      An approximation of $\mathbb{E}_{\pi_n}[h(\theta)]$ is given by

      $$\tilde{h}_{n,J} = \frac{1}{J}\sum_{j=1}^J \tilde{W}_n^j h(\theta_{n-1}^j).$$

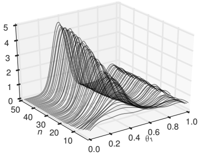
   2. **Selection.**

SMC Algorithm

1. **Initialization.**
2. **Recursion.** For $n = 1, \ldots, N_\phi$,
    1. **Correction.**
    2. **Selection. (Optional Resampling)** Let $\{\hat{\theta}\}_{j=1}^J$ denote J *iid* draws from a multinomial distribution characterized by support points and weights $\{\theta_{n-1}^j, \tilde{W}_n^j\}_{j=1}^J$ and set $W_n^j = 1$.
    An approximation of $\mathbb{E}_{\pi_n}[h(\theta)]$ is given by
    $$\hat{h}_{n,J} = \frac{1}{J}\sum_{j=1}^J W_n^j h(\hat{\theta}_n^j).$$
    3. **Mutation.** Propagate the particles $\{\hat{\theta}_i, W_n^j\}$ via N_{MH} steps of a MH algorithm with transition density $\theta_n^j \sim K_n(\theta_n|\hat{\theta}_n^j; \zeta_n)$ and stationary distribution $\pi_n(\theta)$. An approximation of $\mathbb{E}_{\pi_n}[h(\theta)]$ is given by
    $$\bar{h}_{n,J} = \frac{1}{J}\sum_{j=1}^J h(\theta_n^j) W_n^j.$$

Remarks

- Correction Step:
  - reweight particles from iteration $n-1$ to create importance sampling approximation of $\mathbb{E}_{\pi_n}[h(\theta)]$
- Selection Step: the resampling of the particles
  - (good) equalizes the particle weights and thereby increases accuracy of subsequent importance sampling approximations;
  - (not good) adds a bit of noise to the MC approximation.
- Mutation Step:
  - adapts particles to posterior $\pi_n(\theta)$;
  - imagine we don't do it: then we would be using draws from prior $p(\theta)$ to approximate posterior $\pi(\theta)$, which can't be good!



TV-VARs

- VARs with time-varying parameters (Cogley and Sargent, "Evolving Post-World War II U.S. Inflation Dynamics," NBER MacroAnnual 2001)
- The model:

$$y_t = c_t + \Phi_{1,t} y_{t-1} + \cdots + \Phi_{k,t} y_{t-p} + u_t, \; u_t \sim \mathcal{N}(0, \Sigma), \; t = 1, \ldots, T$$

where y_t and u_t are $n \times 1$ vectors of observables and innovations, c_t is an $n \times 1$ vector of time-varying intercepts,
$\Phi_{1,t}, \ldots, \Phi_{p,t}, \; t = 1, \ldots, T$, and Σ are $n \times n$ matrices.

- We can rewrite the VAR as:

$$y_t = \Phi_t' x_t + u_t$$

  or equivalently as:

$$y_t = X_t' \varphi_t + u_t$$

  where $x_t = [1, y_{t-1}', \ldots, y_{t-p}']'$, $X_t' = I_n \otimes x_t'$, $\Phi_t' = [c_t, \Phi_{1,t}, \ldots, \Phi_{p,t}]$, and $\varphi_t = vec(\Phi_t)$.

- Note $\Phi_t' x_t = vec(x_t' \Phi_t I_n) = (I_n \otimes x_t') vec(\Phi_t)$.

- Assume RW law of motion:

$$\varphi_t = \varphi_{t-1} + \nu_t, \ \nu_t \sim \mathcal{N}(0, Q)$$

with Q being an appropriately-sized positive definite matrix, and

$$\varphi_0 \sim \mathcal{N}(\underline{\phi}, \underline{S}_\phi)$$

where in Cogley and Sargent $\underline{\phi}, \underline{S}_\phi$ are the maximum likelihood mean and variance from pre-sample estimation of a fixed -parameters VAR (alternatively one can use Minnesota prior)

- Belmonte et al. (2011) use the parameterization

$$\varphi_t = \underbrace{\varphi}_{\text{fixed component}} + \underbrace{\tilde{\varphi}_t}_{\text{TV component}}, \ \tilde{\varphi}_0 = 0,$$

(note that under RW φ_0 is not identified; if you assume a stationary law of motion you can identify it), where φ has a Minnesota prior and

$$\tilde{\varphi}_t = \tilde{\varphi}_{t-1} + \nu_t, \ \nu_t \sim \mathcal{N}(0, Q)$$

- Assume independent inverted-Wishart distributions ($\mathcal{IW}(.,.)$) with parameters $(\bar{\Sigma}, \nu_\Sigma)$, $(\bar{Q}, \nu_Q)$, respectively:

$$
\begin{aligned}
p(\Sigma) &= \frac{|\bar{\Sigma}|^{\nu_\Sigma/2}}{2^{n\nu_\Sigma/2}\Gamma(\nu_\Sigma/2)} |\Sigma|^{-(n+\nu_\Sigma+1)/2} \exp\left(-\frac{1}{2}\,\mathrm{tr}(\Sigma^{-1}\bar{\Sigma})\right), \\
p(Q) &= \frac{|\bar{Q}|^{\nu_Q/2}}{2^{n\nu_Q/2}\Gamma(\nu_Q/2)} |Q|^{-(n+\nu_Q+1)/2} \exp\left(-\frac{1}{2}\,\mathrm{tr}(Q^{-1}\bar{Q})\right)
\end{aligned}
$$

TV-VARs: Gibbs Sampler

For $s = 1, \ldots, n_{sim}$:

1. $\varphi_t^{(s)} | Q^{(s-1)}, \Sigma^{(s-1)}, y_{1:T}$: Use the simulation smoother (e.g., in Durbin Koopman 2002), where:

$$y_t = X_t' \varphi_t + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma)$$

is a system of measurement equations and

$$\varphi_t = \varphi_{t-1} + \nu_t, \; \nu_t \sim \mathcal{N}(0, Q)$$

is the system of transition equations.

2. $Q^{(s)} | \varphi_t^{(s)}, \Sigma^{(s-1)}, y_{1:T}$: The product of the law of motion of φ_t and the prior yields:

$$Q^{(s)} | \cdots \sim \mathcal{IW}(\bar{Q} + T \hat{S}_\varphi, \nu_Q + T).$$

where $\hat{S}_\varphi = \dfrac{\sum_{t=1}^{T} (\varphi_t - \varphi_{t-1})(\varphi_t - \varphi_{t-1})'}{T}$.

3. $\Sigma^{(s)}|\varphi_t^{(s)}, Q^{(s)}, y_{1:T}$: The product of the likelihood and the prior yields:

$$\Sigma^{(s)}|\cdots \sim \mathcal{IW}(\bar{\Sigma} + T\hat{S}, \nu_\Sigma + T).$$

where $\hat{S} = \dfrac{\sum_{t=1}^{T}(y_t - X_t'\varphi_t)(y_t - X_t'\varphi_t)'}{T}$.

Stochastic Volatility

- Model (the univariate case):

$$y_t = \sigma_t \varepsilon_t, \ \varepsilon_t \sim \mathcal{N}(0,1)$$

where

$$\sigma_t = e^{\tilde{\sigma}_t}$$

and

$$\tilde{\sigma}_t = \mu + \rho \tilde{\sigma}_{t-1} + \zeta_t, \ \zeta_t \sim \mathcal{N}(0, \omega^2), \ \text{i.i.d. } t,$$

with $\rho < 1$. Call $\theta = \{\mu, \rho, \omega^2\}$.

- If $\rho = 1$ (Primiceri 2005)

$$\tilde{\sigma}_t = \tilde{\sigma}_{t-1} + \zeta_t, \ \zeta_t \sim \mathcal{N}(0, \omega^2), \ \text{i.i.d. } t,$$

with $\tilde{\sigma}_0$ becoming an additional parameter.

Kim, Shephard, Chib (1998)

- Jacquier, Polson, Rossi (1994) provide an alternative approach, but "single move"' (one $\tilde{\sigma}_t$ at the time) $\to$ slow

- Taking squares and then logs of $y_t = \sigma_t \varepsilon_t$, we obtain:

$$e_t^* = 2\tilde{\sigma}_t + \varepsilon_t^*$$

  where $e_t^* = \log(y_t^2 + c)$, $c = .001$ being an offset constant, and $\varepsilon_t^* = \log(\varepsilon_t^2)$.

- If ε_t^* were normally distributed, $\tilde{\sigma}_{1:T}$ could be drawn using standard methods for state-space systems. In fact, ε_t^* is distributed as a $\log(\chi_1^2)$.

- KSC address this problem by approximating the $\log(\chi_1^2)$ with a mixture of normals, that is, expressing the distribution of ε_t^* as:

$$p(\varepsilon_t^*) = \sum_{k=1}^{K} \pi_k^* \mathcal{N}(m_k^* - 1.2704, \nu_k^{*\,2})$$

  The parameters that optimize this approximation, namely $\{\pi_k^*, m_k^*, \nu_k^*\}_{k=1}^K$ and K, are given in KSC for $K = 7$ (or $K = 10$ in Omori, Chib, Shepard, Nakajima JoE 2007). Note that these parameters are independent of the specific application.

- The mixture of normals can be equivalently expressed as:

$$\varepsilon_t^* | \varsigma_t = k \sim \mathcal{N}(m_k^* - 1.2704, \nu_k^{*\,2}), \; Pr(\varsigma_t = k) = \pi_k^*.$$

- Effectively we are replacing the true likelihood $p(y_{1:T}|\theta, \tilde{\sigma}_{1:T})$ with the mixture-of-normal approximation
$\int \tilde{p}(y_{1:T}|\tilde{\sigma}_{1:T}, \theta, \varsigma_{1:T})\pi(\varsigma_{1:T})d\varsigma_{1:T}$

SV: Gibbs sampler

1. $\varsigma_{1:T}^{(s)}|\tilde{\sigma}_{1:T}^{(s-1)}, .., y_{1:T}$: Use

$$Pr\{\varsigma_t = k|\tilde{\sigma}_{1:T}, e^*_{1:T}\} \propto \pi^*_k \nu_k^{-1} \exp\left[-\frac{1}{2\nu_k^{*\,2}}(\varepsilon^*_t - m^*_k + 1.2704)^2\right].$$

where $\varepsilon^*_t = e^*_t - 2\tilde{\sigma}_t$.

2. $\tilde{\sigma}_{1:T}^{(s)}|\varsigma_{1:T}^{(s)}, \theta^{(s-1)}, y_{1:T}$ using

$$e^*_t = 2\tilde{\sigma}_t + m^*_k(\varsigma_t) - 1.2704 + \eta_t,\ \ \eta_t \sim \mathcal{N}(0, \nu^*_k(\varsigma_t)^2)$$

as measurement equations and

$$\tilde{\sigma}_t = \mu + \rho\tilde{\sigma}_{t-1} + \zeta_t,\ \ \zeta_t \sim \mathcal{N}(0, \omega^2),$$

as transition equation.

3. $\theta^{(s)}|\tilde{\sigma}^{(s)}_{1:T}, \varsigma^{(s)}_{1:T}, y_{1:T}$: This is a standard regression problem:

$$\tilde{\sigma}_t = \mu + \rho\tilde{\sigma}_{t-1} + \zeta_t,\ \zeta_t \sim \mathcal{N}(0, \omega^2).$$

- Note that steps 2 and 3 can be integrated in a single block by drawing
$$p(\tilde{\sigma}_{1:T}|\theta, \varsigma_{1:T}, y_{1:T})p(\theta|\varsigma_{1:T}, y_{1:T})$$
where
    - $\tilde{\sigma}_{1:T}$ are integrated out using the Kalman filter $\longrightarrow$ θ is drawn from $p(\theta|\varsigma_{1:T}, y_{1:T})$ using MH.
    - $p(\tilde{\sigma}_{1:T}|\theta, \varsigma_{1:T}, y_{1:T})$ are drawn using the simulation smoother

An Example of a Wrong Gibbs Sampler

- This is an example from Del Negro, Primiceri (2013) "Time Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum", which corrects a mistake in Primiceri (2005)
- Take the model of Kim, Shepard, Chib (1998), **except for the constant** θ:

$$y_t = \theta + \sigma_t \varepsilon_t, \ \varepsilon_t \sim \mathcal{N}(0,1)$$

where y_t is univariate and

$$\sigma_t = e^{\tilde{\sigma}_t}$$

and

$$\tilde{\sigma}_t = \tilde{\sigma}_{t-1} + \zeta_t, \ \zeta_t \sim \mathcal{N}(0, \omega^2), \text{ i.i.d. } t$$

- Assume you know ω^2 and the initial condition $\tilde{\sigma}_0$ for simplicity.

Primiceri's Gibbs Sampler

- This is a **three-block** Sampler in $\tilde{\sigma}_{1:T},\ \varsigma_{1:T},\ \theta$

1. **Draw $\tilde{\sigma}_{1:T}$ from**
$\tilde{p}\left(\tilde{\sigma}_{1:T}|y_{1:T},\theta,\varsigma_{1:T}\right) \propto \tilde{p}(y_{1:T}|\tilde{\sigma}_{1:T},\theta,\varsigma_{1:T}) \cdot p(\tilde{\sigma}_{1:T})$. Specifically, use

$$e_t^* = 2\tilde{\sigma}_t + m_k^*(s_t) - 1.2704 + \eta_t,\ \eta_t \sim \mathcal{N}(0, \nu_k^*(s_t)^2),$$

where $e_t^* = \log\left((y_t - \theta)^2 + c\right)$, $c = .001$ being an offset constant, and $\varepsilon_t^* = \log(\varepsilon_t^2)$, as measurement equations and

$$\tilde{\sigma}_t = \tilde{\sigma}_{t-1} + \zeta_t,\ \zeta_t \sim \mathcal{N}(0, \omega^2),$$

as transition equation

- Use simulation smoothers (Carter and Kohn 1994, Durbin and Koopman 2001)

2. **Draw** $\varsigma_{1:T}$ **from** $\tilde{p}\left(\varsigma_{1:T}|y_{1:T}, \tilde{\sigma}_{1:T}, \theta\right) \propto \tilde{p}(y_{1:T}|\Sigma^T, \theta, \varsigma_{1:T}) \cdot \pi(\varsigma_{1:T})$. Specifically, use

$$Pr\{s_t = k|\tilde{\sigma}_{1:T}, e^*_{1:T}\} \propto \pi^*_k \nu_k^{-1} \exp\left[-\frac{1}{2\nu_k^{*\,2}}(\varepsilon^*_t - m^*_k + 1.2704)^2\right].$$

where $\varepsilon^*_t = e^*_t - 2\tilde{\sigma}_t$.

3. **Draw** θ **from** $p\left(\theta|y_{1:T}, \tilde{\sigma}_{1:T}\right) \propto p(y_{1:T}|\tilde{\sigma}_{1:T}, \theta) \cdot p(\theta)$. Standard GLS regression:

$$y_t = \theta + \sigma_t \varepsilon_t, \ \varepsilon_t \sim \mathcal{N}(0, 1)$$

Two Problems with Primiceri's Gibbs Sampler

1. Steps (1) and (2) use the approximate likelihood $\tilde{p}(.)$, step (3) uses the true likelihood $p(.)$
   - KSC do not have step (3) $\rightarrow$ they only use $\tilde{p}(.)$ $\rightarrow$ they can address the approximation issue by re-weighting all draws by the ratio of true vs approximate likelihood *at the end of the sampler* (re-weigthing usually makes little difference)
   - Step (3) prevents us from using this fix

2. This is **not** a correct three-block sampler!
   1. Draw $\tilde{\sigma}_{1:T}$ from $\tilde{p}\left(\tilde{\sigma}_{1:T} | y_{1:T}, \theta, \varsigma_{1:T}\right)$
   2. Draw $\varsigma_{1:T}$ from $\tilde{p}\left(\varsigma_{1:T} | y_{1:T}, \tilde{\sigma}_{1:T}, \theta\right)$
   3. Draw θ from $p\left(\theta | y_{1:T}, \tilde{\sigma}_{1:T}, ???\right)$

- Using $p\left(\theta|y_{1:T},\tilde{\sigma}_{1:T},\varsigma_{\mathbf{1:T}}\right)$ in step (3) is **not** a (convenient) **solution** in macro models: Conditional on $\varsigma_{1:T}$, $\varepsilon_t^* = \log(\varepsilon_t^2)$ is Gaussian, but this means that ε_t is no longer Gaussian in $y_t = \theta + \sigma_t\varepsilon_t$

A Solution to Problem # 2

- Assume *for now* that the mixture-of-normal distribution is correct:

$$p(y_{1:T}|\theta, \tilde{\sigma}_{1:T}) = \int \tilde{p}(y_{1:T}|\tilde{\sigma}_{1:T}, \theta, \varsigma_{1:T})\pi(\varsigma_{1:T})d\varsigma_{1:T} \tag{1}$$

- Say you know how to obtain draws $\{\theta^{(j)}, \Sigma^{T\ (j)}, s^{T\ (j)}\}_{j=1}^{n_s}$ from the joint

$$\tilde{p}(\tilde{\sigma}_{1:T}, \theta, \varsigma_{1:T}|y_{1:T}) = \tilde{p}(y_{1:T}|\Sigma^T, \theta, \varsigma_{1:T}) \cdot p(\tilde{\sigma}_{1:T}, \theta) \cdot \pi(\varsigma_{1:T})$$

- Then the draws $\{\theta^{(j)}, \Sigma^{T\ (j)}\}_{j=1}^{n_s}$ obtained this way are what we want since (1) implies

$$\int \tilde{p}(\tilde{\sigma}_{1:T}, \theta, \varsigma_{1:T}|y_{1:T})d\varsigma_{1:T} = p(y_{1:T}|\Sigma^T, \theta) \cdot p(\tilde{\sigma}_{1:T}, \theta)$$

A Solution to Problem # 2

- We can draw from the joint $\tilde{p}(\tilde{\sigma}_{1:T}, \theta, \varsigma_{1:T}|y_{1:T})$ using the following sampler

  1. Draw $\tilde{\sigma}_{1:T}$ from
     $\tilde{p}\left(\Sigma^T|y_{1:T}, \theta, \varsigma_{1:T}\right) \propto \tilde{p}(y_{1:T}|\tilde{\sigma}_{1:T}, \theta, \varsigma_{1:T}) \cdot p(\tilde{\sigma}_{1:T}|\theta)$
  2. Draw $(\theta, \varsigma_{1:T})$ from $\tilde{p}\left(\theta, \varsigma_{1:T}|y_{1:T}, \tilde{\sigma}_{1:T}\right)$, which is accomplished by

     (i) Drawing θ from the **marginal**
     $p\left(\theta|y_{1:T}, \tilde{\sigma}_{1:T}\right) \propto p(y_{1:T}|\tilde{\sigma}_{1:T}, \theta) \cdot p(\theta|\tilde{\sigma}_{1:T})$.

     (ii) Drawing $\varsigma_{1:T}$ from the **conditional**
     $\tilde{p}\left(\varsigma_{1:T}|y_{1:T}, \tilde{\sigma}_{1:T}, \theta\right) \propto \tilde{p}(y_{1:T}|\Sigma^T, \theta, \varsigma_{1:T}) \cdot \pi(\varsigma_{1:T})$.

- In step (2.i) we are entitled to use $p(.)$ (and not $\tilde{p}(.)$) because we have integrated out the $\varsigma_{1:T}$ and
  $$p(y_{1:T}|\theta, \tilde{\sigma}_{1:T}) = \int \tilde{p}(y_{1:T}|\tilde{\sigma}_{1:T}, \theta, \varsigma_{1:T})\pi(\varsigma_{1:T})d\varsigma_{1:T}$$
- These are exactly the same steps as in Primiceri, **but we need to draw $\varsigma_{1:T}$ right before $\tilde{\sigma}_{1:T}$!**

A Solution to Problem # 1 (Approximation)

- As long as the number of components in the mixture is large enough (10?) this is not a big deal.
- Stroud, Müller and Polson (2003) show how to fix it
- Construct a joint posterior of $\tilde{\sigma}_{1:T}$, θ and $\varsigma_{1:T}$ as follows:

$$\begin{aligned} p(\theta, \tilde{\sigma}_{1:T}, \varsigma_{1:T} | y_{1:T}) &= p(\theta, \tilde{\sigma}_{1:T} | y_{1:T}) \cdot \tilde{p}(\varsigma_{1:T} | \tilde{\sigma}_{1:T}, \theta, y_{1:T}) \\ &\propto \underbrace{p(y_{1:T} | \theta, \tilde{\sigma}_{1:T}) \cdot p(\tilde{\sigma}_{1:T}, \theta)}_{\text{original posterior}} \cdot \tilde{p}(\varsigma_{1:T} | \tilde{\sigma}_{1:T}, \theta, y_{1:T}), \end{aligned} \tag{2}$$

with

$$\tilde{p}(\varsigma_{1:T} | \tilde{\sigma}_{1:T}, \theta, y_{1:T}) = \frac{\tilde{p}(y_{1:T} | \Sigma^T, \theta, \varsigma_{1:T}) \cdot \pi(\varsigma_{1:T})}{c(\tilde{\sigma}_{1:T}, \theta, y_{1:T})}, \tag{3}$$

where $c(\tilde{\sigma}_{1:T}, \theta, y_{1:T}) \equiv \int \tilde{p}(y_{1:T} | \Sigma^T, \theta, \varsigma_{1:T}) \pi(\varsigma_{1:T}) d\varsigma_{1:T}$ guarantees that the density in (3) integrates to one.
- Obviously drawing from (2) yields the correct draws

The Correct Algorithm

1. Draw $\tilde{\sigma}_{1:T}$ from $p\left(\tilde{\sigma}_{1:T}|y_{1:T},\theta,\varsigma_{1:T}\right)$ as follows: Draw a candidate $\tilde{\sigma}^{\dagger}_{1:T}$ from the proposal density $\tilde{p}\left(\tilde{\sigma}_{1:T}|y_{1:T},\theta,\varsigma_{1:T}\right)$ of Algorithm 2, and set

$$\tilde{\sigma}^{(j)}_{1:T} = \begin{cases} \tilde{\sigma}^{\dagger}_{1:T} & \text{with probability } \alpha \\ \tilde{\sigma}^{(j-1)}_{1:T} & \text{with probability } 1-\alpha \end{cases},$$

where the superscript (j) denotes the iteration of the sampler, and where

$$\alpha = \frac{p\left(\tilde{\sigma}^{\dagger}_{1:T}|y_{1:T},\theta,\varsigma_{1:T}\right)}{p\left(\tilde{\sigma}^{(j-1)}_{1:T}|y_{1:T},\theta,\varsigma_{1:T}\right)} \frac{\tilde{p}\left(\tilde{\sigma}^{(j-1)}_{1:T}|y_{1:T},\theta,\varsigma_{1:T}\right)}{\tilde{p}\left(\tilde{\sigma}^{\dagger}_{1:T}|y_{1:T},\theta,\varsigma_{1:T}\right)}.$$

- The acceptance probability can be rewritten as

$$\alpha = \frac{p(y_{1:T}|\theta,\tilde{\sigma}^{\dagger}_{1:T})}{p(y_{1:T}|\theta,\Sigma^{(j-1)\ T})} \frac{c(\tilde{\sigma}^{(j-1)}_{1:T},\theta,y_{1:T})}{c(\tilde{\sigma}^{\dagger}_{1:T},\theta,y_{1:T})}.$$

where where $c(\tilde{\sigma}_{1:T},\theta,y_{1:T}) \equiv \int \tilde{p}(y_{1:T}|\Sigma^{T},\theta,\varsigma_{1:T})\pi(\varsigma_{1:T})d\varsigma_{1:T}$ is precisely the mixture-of-normal approximation!

The Correct Algorithm

2. Draw $(\theta, \varsigma_{1:T})$ from $p\left(\theta, \varsigma_{1:T} | y_{1:T}, \Sigma^T\right)$, which is accomplished by

   (i) Drawing θ from

$$
\begin{aligned}
p\left(\theta | y_{1:T}, \tilde{\sigma}_{1:T}\right) &= \int p(\theta, \varsigma_{1:T} | y_{1:T}, \tilde{\sigma}_{1:T}) d\varsigma_{1:T} \\
&\propto p(y_{1:T} | \theta, \tilde{\sigma}_{1:T}) \cdot p(\theta | \tilde{\sigma}_{1:T}) \cdot \int \tilde{p}(\varsigma_{1:T} | \tilde{\sigma}_{1:T}, \theta, y_{1:T} \\
&= p(y_{1:T} | \tilde{\sigma}_{1:T}, \theta) \cdot p(\theta | \tilde{\sigma}_{1:T}).
\end{aligned}
$$

   (ii) Drawing $\varsigma_{1:T}$ from
   $\tilde{p}\left(\varsigma_{1:T} | y_{1:T}, \tilde{\sigma}_{1:T}, \theta\right) \propto \tilde{p}(y_{1:T} | \Sigma^T, \theta, \varsigma_{1:T}) \cdot \pi(\varsigma_{1:T})$.

Affected Applications

- Primiceri (2005)'s TV-VAR with SV:

$$y_t = c_t + B_{1,t}y_{t-1} + ... + B_{k,t}y_{t-k} + A_t^{-1}\Sigma_t\varepsilon_t$$

where all the TV coefficients evolve as random walks, and all the innovations in the model are jointly normally distributed with covariance matrix equal to V. Define $\theta \equiv \left[B^T, A^T, V\right]$

- Stock and Watson (2007)'s unobserved component model with SV:

$$y_t = c_t + \sigma_{\varepsilon,t}\varepsilon_t$$

where

$$c_t = c_{t-1} + \sigma_{e,t}e_t$$

Define $\theta \equiv \left[c^T\right]$

- Del Negro and Otrok (2008)'s factor model with SV:

$$y_{i,t} = a_{i,t} + \lambda_{i,t} f_t + \xi_{i,t}, \quad t = 1, \ldots, T.$$

where

$$f_t = \Phi_{0,1} f_{t-1} + \ldots + \Phi_{0,q} f_{t-q} + u_{0,t}, \quad u_{0,t} \sim iidN(0, \Sigma_{0,t}),$$

and

$$\xi_{i,t} = \phi_{i,1} \xi_{i,t-1} + \ldots + \phi_{i,p_i} \xi_{i,t-p_i} + u_{i,t}, \quad u_{i,t} \sim iidN(0, \sigma^2_{i,t}).$$

Define $\theta \equiv \left[A^T, \Lambda^T, f^T, \Phi\right]$

- DSGEs with SV (Justiniano, Primiceri (2008) and Cúrdia, Del Negro, Greenwald (2014))

Dynamic Factor Models

- A DFM decomposes the dynamics of n observables $y_{i,t}$, $i = 1, \ldots, n$, into the sum of two unobservable components:

$$y_{i,t} = a_i + \lambda_i f_t + \xi_{i,t}, \quad t = 1, \ldots, T.$$

- Here f_t is a $\kappa \times 1$ vector of factors that are common to all observables and $\xi_{i,t}$ is an idiosyncratic process, that is specific to each i.
- The factors follow a vector autoregressive processes of order q:

$$f_t = \Phi_{0,1} f_{t-1} + \ldots + \Phi_{0,q} f_{t-q} + u_{0,t}, \quad u_{0,t} \sim iidN(0, \Sigma_0),$$

- The idiosyncratic components follow autoregressive processes of order p_i:

$$\xi_{i,t} = \phi_{i,1}\xi_{i,t-1} + \ldots + \phi_{i,p_i}\xi_{i,t-p_i} + u_{i,t}, \quad u_{i,t} \sim iidN(0, \sigma_i^2).$$

Identification

- Without further restrictions the latent factors and the coefficient matrices of the DFM are not identifiable.

- One can premultiply f_t as well as $u_{0,t}$ by a $\kappa \times \kappa$ invertible matrix H and post-multiply the vectors λ_i and the matrices $\Phi_{0,j}$ by H^{-1}, without changing the distribution of the observables.

- We provide three examples of achieving identification.

Example 1 – Geweke and Zhou (1996)

- Restrict $\Lambda_{1,\kappa}$ to be lower triangular:

$$\Lambda_{1,\kappa} = \Lambda^{tr}_{1,\kappa} = \begin{bmatrix} X & 0 \cdots 0 & 0 \\ \vdots & \ddots & \vdots \\ X & X \cdots X & X \end{bmatrix}.$$

- Factors and hence matrices $\Phi_{0,j}$ and Σ_0 could still be transformed by an arbitrary invertible lower triangular $\kappa \times \kappa$ matrix H_{tr} without changing the distribution of the observables.
- Under this transformation the factor innovations become $H_{tr} u_{0,t}$.
- Choose $H_{tr} = \Sigma^{-1}_{0,tr}$ such that the factor innovations reduce to a vector of independent standard Normals.
- To implement this normalization, we simply let

$$\Sigma_0 = I_\kappa.$$

- Sign normalization can be achieved with a set of restrictions of the form:

$$\lambda_{i,i} \geq 0, \quad i = 1, \ldots, \kappa.$$

Example 2

- $\Lambda_{1,\kappa}$ is restricted to be lower triangular with ones on the diagonal and Σ_0 is a diagonal matrix with non-negative elements.
- The one-entries on the diagonal of $\Lambda_{1,\kappa}$ also take care of the sign normalization.
- Since under the normalization $\lambda_{i,i} = 1$, $i = 1, \ldots, \kappa$, factor $f_{i,t}$ is forced to have a unit impact on $y_{i,t}$

Example 3

- $\Lambda_{1,\kappa}$ is restricted to be the identity matrix and Σ_0 is an unrestricted covariance matrix.
- The one-entries on the diagonal of $\Lambda_{1,\kappa}$ take care of the sign normalization.

Joint Distribution – Assume $p_i = p$, $q \leq p+1$

Quasi-differenced Measurement Equation:

$$\begin{aligned} y_{i,t} &= a_i + \lambda_i f_t + \phi_{i,1}(y_{i,t-1} - a_i - \lambda_i f_{t-1}) + \ldots \\ &\quad + \phi_{i,p}(y_{i,t-p} - a_i - \lambda_i f_{t-p}) + u_{i,t}, \quad \text{for} \quad t = p+1, .., T. \end{aligned}$$

Joint distribution:

$$\begin{aligned} & p(y_{1:T}, f_{0:T}, \{\theta_i\}_{i=1}^n, \theta_0) \\ &= \left[\prod_{t=p+1}^{T} \left(\prod_{i=1}^{n} p(y_{i,t} | y_{i,t-p:t-1}, f_{t-p:t}, \theta_i) \right) p(f_t | f_{t-q:t-1}, \theta_0) \right] \\ &\quad \times \left(\prod_{i=1}^{n} p(y_{i,1:p} | f_{0:p}, \theta_i) \right) p(f_{0:p} | \theta_0) \left(\prod_{i=1}^{n} p(\theta_i) \right) p(\theta_0). \end{aligned}$$

where θ_0 determines the law of motion of the factors and θ_i summarizes unit-specific coefficients.

Priors are conjugate.

Gibbs Sampler: $\theta_i|\cdot$

The posterior density takes the form:

$$p(\theta_i|f_{0:T},\theta_0,y_{1:T}) \quad \propto \quad p(\theta_i)\left(\prod_{t=p+1}^{T} p(y_{i,t}|y_{i,t-p:t-1},f_{t-p:t},\theta_i)\right)$$
$$p(y_{i,1:p}|f_{0:p},\theta_i).$$

- Use Chib and Greenberg (1994)'s procedure to generate draws from a regression with $AR(p)$ errors (see Otrok and Whiteman, 1998 and these notes -*caveat emptor*!- of mine)
- If prior for $\lambda_{i,i}$, $i=1,\ldots,\kappa$ includes $\mathcal{I}\{\lambda_{i,i}\geq 0\}$, one can use an acceptance sampler that discards all draws of θ_i for which $\lambda_{i,i}<0$.
- If the prior is symmetric around zero, then one can resolve the sign indeterminacy by post-processing the output of the (unrestricted) Gibbs sampler: for each set of draws $(\{\theta_i\}_{i=1}^n,\theta_0,f_{0:T})$ such that $\lambda_{i,i}<0$, flip the sign of the i'th factor and the sign of the loadings of all n observables on the ith factor.

Gibbs Sampler: $\theta_0|\cdot$

The posterior density takes the form:

$$p(\theta_0|f_{0:T}, \{\theta_i\}_{i=1}^n, y_{1:T}) \quad \propto \quad \left(\prod_{t=p+1}^{T} p(f_t|F_{t-p,t-1}, \theta_0)\right)$$
$$p(\theta_0)p(f_{0:p}|\theta_0)$$

Gibbs Sampler: $f_{0:T}|\cdot$

- Write DFM in state-space form...
- Measurement equation (stack measurement eqs for all is):

$$(I_n - \sum_{j=1}^{p} \tilde{\Phi}_j L^j) y_t = (I_n - \sum_{j=1}^{p} \tilde{\Phi}_j) a + \Lambda^* \tilde{f}_t + u_t, \quad t = p+1, \dots, T,$$

where $y_t = [y_{1,t}, .., y_{n,t}]'$, $a_t = [a_1, .., a_n]'$, and $u_t = [u_{1,t}, .., u_{n,t}]'$, $\tilde{\Phi}_j$s are diagonal matrices with $[\phi_{1,j}, .., \phi_{n,j}]'$ on the diagonal, $\tilde{f}_t = [f_t', .., f_{t-p}']'$, and

$$\Lambda^* = \begin{bmatrix} \lambda_1 & -\lambda_1\phi_{1,1} & \dots & -\lambda_1\phi_{1,p} \\ \vdots & & \ddots & \vdots \\ \lambda_n & -\lambda_n\phi_{n,1} & \dots & -\lambda_n\phi_{n,p} \end{bmatrix}$$

- Note: u_t is *iid*!

- Factor law of motion in companion form (transition equation):

$$\tilde{f}_t = \tilde{\Phi}_0 \tilde{f}_{t-1} + \tilde{u}_{0,t},$$

where

$$\tilde{\Phi}_0 = \begin{bmatrix} \Phi_{0,1} & \Phi_{0,2} & \cdots & \Phi_{0,p} & 0_{k\times k(p+1-q)} \\ I & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}$$

- Since the measurement equation starts from $t = p + 1$ as opposed to $t = 1$, one needs to initialize the filtering step in the Carter-Kohn algorithm with the conditional distribution of $p(f_{0:p}|y_{1:p}, \{\theta_i\}_{i=1}^n, \theta_0)$.

Factor Model: Gibbs Sampler

For $s = 1, \dots, n_{sim}$:

1. Draw $\theta_i^{(s)}$ conditional on $(f_{0:T}^{(s-1)}, \theta_0^{(s-1)}, y_{1:T})$. This can be done independently for each $i = 1, \dots, n$.

2. Draw $\theta_0^{(s)}$ conditional on $(f_{0:T}^{(s-1)}, \{\theta_i^{(s)}\}_{i=1}^n, y_{1:T})$.

3. Draw $f_{0:T}^{(s)}$, conditional on $(\{\theta_i^{(s)}\}_{i=1}^n, \theta_0^{(s)}, y_{1:T})$.

- Here are the codes for the Del Negro Otrok paper ("99 Luftballons: Monetary policy and the house price boom across US states", JME 2007)

Factor Augmented VARs

- FAVARs allow for additional observables $y_{0,t}$, e.g., the Fed Funds rate, to enter the measurement equation, which becomes:

$$y_{i,t} = a_i + \gamma_i y_{0,t} + \lambda_i f_t + \xi_{i,t}, \quad i = 1, \dots, n, \quad t = 1, \dots, T,$$

  where $y_{0,t}$ and γ_i are $m \times 1$ and $1 \times m$ vectors, respectively.

- The observable vector $y_{0,t}$ and the unobservable factor f_t are assumed to jointly follow a vector autoregressive processes of order q:

$$\begin{bmatrix} f_t \\ y_{0,t} \end{bmatrix} = \Phi_{0,1} \begin{bmatrix} f_{t-1} \\ y_{0,t-1} \end{bmatrix} + \dots + \Phi_{0,q} \begin{bmatrix} f_{t-q} \\ y_{0,t-q} \end{bmatrix} + u_{0,t},$$

  $u_{0,t} \sim iidN(0, \Sigma_0)$ which is the reason for the term factor augmented VAR.

- Identification:

$$u_{0,t} = \Sigma_{0,tr} \Omega_0 \epsilon_{0,t}.$$

Linear DSGEs

- A simple DSGE model
- Parameters estimation
- Impulse responses and variance decomposition
- Inference on latent variables: Shock decomposition

A Prototypical DSGE Model: Household

- Preferences:

$$\boldsymbol{E}_t \left[\sum_{s=0}^{\infty} \beta^{t+s} \left(\ln C_{t+s} - \frac{(H_{t+s}/B_{t+s})^{1+1/\nu}}{1+1/\nu} \right) \right]$$

- Budget constraint:

$$C_t + I_t \leq W_t H_t + R_t K_t.$$

- Capital accumulation:

$$K_{t+1} = (1-\delta)K_t + I_t,$$

- First-order conditions:

$$\frac{1}{C_t} = \beta \boldsymbol{E} \left[\frac{1}{C_{t+1}} (R_{t+1} + (1-\delta)) \right] \quad \text{and} \quad \frac{1}{C_t} W_t = \frac{1}{B_t} \left(\frac{H_t}{B_t} \right)^{1/\nu}.$$

Firms

- Technology:

$$Y_t = (A_t H_t)^{\alpha} K_t^{1-\alpha}.$$

- First-order conditions from profit maximization:

$$W_t = \alpha \frac{Y_t}{H_t}, \quad R_t = (1-\alpha)\frac{Y_t}{K_t}.$$

- Market clearing:

$$Y_t = C_t + I_t.$$

Exogenous Processes

- Log Technology:

$$\ln A_t = \ln A_0 + (\ln \gamma)t + \ln \widetilde{A}_t, \quad \ln \widetilde{A}_t = \rho_a \ln \widetilde{A}_{t-1} + \sigma_a \epsilon_{a,t}$$

where $\epsilon_{a,t} \sim iidN(0,1)$.

- Preference shifter:

$$\ln B_t = (1 - \rho_b) \ln B_* + \rho_b \ln B_{t-1} + \sigma_b \epsilon_{b,t}$$

where $\epsilon_{b,t} \sim iidN(0,1)$.

- Initialzation:

$$\ln \widetilde{A}_{-\tau} = 0 \quad \text{and} \quad \ln B_{-\tau} = 0.$$

Model Solution

- The solution to the rational expectations difference equations determines law of motion for Y_t, C_t, I_t, K_t, H_t, W_t, and R_t.
- The technology process $\ln A_t$ induces a common trend in output, consumption, investment, capital, and wages.
- It is useful to detrended the model variables as follows:

$$\widetilde{Y}_t = \frac{Y_t}{A_t},\ \widetilde{C}_t = \frac{C_t}{A_t},\ \widetilde{I}_t = \frac{I_t}{A_t},\ \widetilde{K}_{t+1} = \frac{K_{t+1}}{A_t},\ \widetilde{W}_t = \frac{W_t}{A_t}.$$

Equilibrium Conditions, Rewritten

$$
\begin{aligned}
\frac{1}{\widetilde{C}_t} &= \beta \boldsymbol{E}\left[\frac{1}{\widetilde{C}_{t+1}} e^{-a_{t+1}}(R_{t+1} + (1-\delta))\right], \quad \frac{1}{\widetilde{C}_t}\widetilde{W}_t = \frac{1}{B_t}\left(\frac{H_t}{B_t}\right)^{1/\nu} \\
\widetilde{W}_t &= \alpha \frac{\widetilde{Y}_t}{H_t}, \quad R_t = (1-\alpha)\frac{\widetilde{Y}_t}{\widetilde{K}_t} e^{a_t} \\
\widetilde{Y}_t &= H_t^{\alpha}\left(\widetilde{K}_t e^{-a_t}\right)^{1-\alpha}, \quad \widetilde{Y}_t = \widetilde{C}_t + \widetilde{I}_t, \quad \widetilde{K}_{t+1} = (1-\delta)\widetilde{K}_t e^{-a_t} + \widetilde{I}_t.
\end{aligned}
$$

- The process a_t is defined as

$$a_t = \ln \frac{A_t}{A_{t-1}} = \ln \gamma + (\rho_a - 1)\ln \widetilde{A}_{t-1} + \sigma_a \epsilon_{a,t}.$$

- This log ratio is always stationary, because if $\rho_a = 1$ the $\ln \widetilde{A}_{t-1}$ term drops out.

Steady State, Etc.

- Steady state:

$$R_* = \frac{\gamma}{\beta} - (1-\delta), \quad \frac{\widetilde{K}_*}{\widetilde{Y}_*} = \frac{(1-\alpha)\gamma}{R_*}, \quad \frac{\widetilde{I}_*}{\widetilde{Y}_*} = \left(1 - \frac{1-\delta}{\gamma}\right)\frac{\widetilde{K}_*}{\widetilde{Y}_*}.$$

- If $\rho_a = 1$, the model generates cointegration relationships which are obtained by taking pair-wise differences of $\ln Y_t$, $\ln C_t$, $\ln I_t$, $\ln K_{t+1}$, and $\ln W_t$
- Parameters are stacked in vector θ:

$$\theta = [\alpha, \beta, \gamma, \delta, \nu, \rho_a, \sigma_a, \rho_b, \sigma_b]'.$$

Loglinearization

$$
\begin{aligned}
\widehat{C}_t &= \boldsymbol{E}_t\left[\widehat{C}_{t+1} + \widehat{a}_{t+1} - \frac{R_*}{R_* + (1-\delta)}\widehat{R}_{t+1}\right] \\
\widehat{H}_t &= \nu\widehat{W}_t - \nu\widehat{C}_t + (1+\nu)\widehat{B}_t, \quad \widehat{W}_t = \widehat{Y}_t - \widehat{H}_t, \\
\widehat{R}_t &= \widehat{Y}_t - \widehat{K}_t + \widehat{a}_t, \quad \widehat{K}_{t+1} = \frac{1-\delta}{\gamma}\widehat{K}_t + \frac{\widetilde{I}_*}{\widetilde{K}_*}\widehat{I}_t - \frac{1-\delta}{\gamma}\widehat{a}_t, \\
\widehat{Y}_t &= \alpha\widehat{H}_t + (1-\alpha)\widehat{K}_t - (1-\alpha)\widehat{a}_t, \quad \widehat{Y}_t = \frac{\widetilde{C}_*}{\widetilde{Y}_*}\widehat{C}_t + \frac{\widetilde{I}_*}{\widetilde{Y}_*}\widehat{I}_t, \\
\widehat{A}_t &= \rho_a\widehat{A}_{t-1} + \sigma_a\epsilon_{a,t}, \quad \widehat{a}_t = \widehat{A}_t - \widehat{A}_{t-1}, \quad \widehat{B}_t = \rho_b\widehat{B}_{t-1} + \sigma_b\epsilon_{b,t}.
\end{aligned}
$$

Log-linearization of $f(x)$:

1. write $f(x) = f(e^z)$;
2. conduct a first-order Taylor approximation around x_0 in terms of z:

$$f(e^{\ln x}) \approx f(x_0) + x_0 f^{(1)}(x_0)(\ln x - \ln x_0).$$

What have we got? A state-space model!

Model solution

- Need to solve for expectations!
- We can follow the method in Sims (2002) (Christopher A. Sims, "Solving Linear Rational Expectations Models". Computational Economics, Vol. 20 (1-2), 1-20), implemented using the code `gensys` (available in Matlab and R on Chris' webpage).
- For any *endogenous* variable x_t for which $E_t[x_{t+1}]$ appears in the equilibrium conditions (e.g., $E_t[\hat{\pi}_{t+1}]$) define the variable $x_t^E = E_t[x_{t+1}]$ (e.g., $\hat{\pi}_{t+1}^E = E_t[\hat{\pi}_{t+1}]$) and note that

$$\begin{aligned} x_t &= E_{t-1}[x_t] + \eta_{x,t} \\ &= x_{t-1}^E + \eta_{x,t} \end{aligned}$$

where rational expectations implies that:

$$E_t[\eta_{x,t+1}] = 0$$

- Write the system (equilibrium conditions, evolution exogenous processes, expectational equations) as:

$$\Gamma_0 s_t = \Gamma_1 s_{t-1} + \Psi \varepsilon_t + \Pi \eta_t$$

where:

1. s_t is a vector including all endogenous, exogenous variables + expectational terms (e.g. $s_t = \{\hat{\pi}_t, .., \pi_t^E, .., z_t, ..\}$)
2. ε_t includes all innovations to exogenous processes (e.g. $\varepsilon_t = \{\varepsilon_{z,t}, \varepsilon_{g,t}, ...\}$)
3. η_t includes all expectational errors (e.g. $\eta_t = \{\eta_{\pi,t}, ...\}$).

- $$\Gamma_0(\theta)s_t = \Gamma_1(\theta)s_{t-1} + \Psi(\theta)\varepsilon_t + \Pi(\theta)\eta_t$$

$$\Downarrow$$

$$\texttt{gensys}$$

$$\Downarrow$$

$$s_t = T(\theta)s_{t-1} + R(\theta)\varepsilon_t$$

Measurement equation: Examples

- Observations on GDP and Hours:

$$\begin{bmatrix} \ln GDP_t \\ \ln H_t \end{bmatrix} = \begin{bmatrix} \ln Y_0 \\ \ln H_* \end{bmatrix} + \begin{bmatrix} \ln \gamma \\ 0 \end{bmatrix} t + \begin{bmatrix} \widehat{Y}_t + \widehat{A}_t \\ \widehat{H}_t \end{bmatrix},$$

- Observations on GDP and Investment:

$$\begin{bmatrix} \ln GDP_t \\ \ln I_t \end{bmatrix} = \begin{bmatrix} \ln Y_0 \\ \ln Y_0 + (\ln \widetilde{I}_* - \ln \widetilde{Y}_*) \end{bmatrix} + \begin{bmatrix} \ln \gamma \\ \ln \gamma \end{bmatrix} t + \begin{bmatrix} \widehat{A}_t + \widehat{Y}_t \\ \widehat{A}_t + \widehat{I}_t \end{bmatrix}.$$

Bayesian Estimation – Prior

Name	Domain	Density	Para (1)	Para (2)
α	$[0,1)$	Beta	0.66	0.02
ν	$I\!R^+$	Gamma	2.00	1.00
$4\ln\gamma$	$I\!R$	Normal	0.00	0.10
ρ_a	$I\!R^+$	Beta	0.95	0.02
σ_a	$I\!R^+$	InvGamma	0.01	4.00
ρ_b	$I\!R^+$	Beta	0.80	0.10
σ_b	$I\!R^+$	InvGamma	0.01	4.00
$\ln H_*$	$I\!R$	Normal	0.00	10.0
$\ln Y_0$	$I\!R$	Normal	0.00	100

Estimation: Random-Walk Metropolis Algorithm for DSGE Model

1. Construct the proposal density:

   1.a Use a numerical optimization routine to maximize the log posterior: $\ln p(y_{1:T}|\theta) + \ln p(\theta)$. Call $\tilde{\theta}$ the posterior mode.

   1.b Compute numerically the inverse of the (negative) Hessian computed at the posterior mode $\tilde{\theta}$, call it $\tilde{\Sigma}$.

2. Draw $\theta^{(0)}$ from $N(\tilde{\theta}, c_0^2\tilde{\Sigma})$ or directly specify a starting value.

3. For $j = 1, \dots, n_{sim}$: draw θ^* from the proposal distribution $N(\theta^{(j-1)}, c^2\tilde{\Sigma})$. The jump from $\theta^{(j-1)}$ is accepted ($\theta^{(j)} = \theta^*$) with probability $\min\{1, r(\theta^{(j-1)}, \theta^*|y_{1:T})\}$ and rejected ($\theta^{(j)} = \theta^{(j-1)}$) otherwise, where

$$r(\theta^{(j-1)}, \theta^*|y_{1:T}) = \frac{p(y_{1:T}|\theta^*)p(\theta^*)}{p(y_{1:T}|\theta^{(j-1)})p(\theta^{(j-1)})}$$

4. Burn-in period: throw away draws $\theta^{(j)}$, for $j = 1, .., n^{burn}$, where $n^{burn}/n^{sim} \approx 10\%$.

- Matlab estimation code for the FRBNY DSGE model

Posterior (simple RBC example)

	Det. Trend		Stoch. Trend	
Name	Mean	90% Intv.	Mean	90% Intv.
α	0.65	[0.62, 0.68]	0.65	[0.63, 0.69]
ν	0.42	[0.16, 0.67]	0.70	[0.22, 1.23]
$4 \ln \gamma$	.003	[.002, .004]	.004	[.002, .005]
ρ_a	0.97	[0.95, 0.98]	1.00	
σ_a	.011	[.010, .012]	.011	[.010, .012]
ρ_b	0.98	[0.96, 0.99]	0.98	[0.96, 0.99]
σ_b	.008	[.007, .008]	.007	[.006, .008]
$\ln H_*$	-0.04	[-0.08, 0.01]	-0.03	[-0.07, 0.02]
$\ln Y_0$	8.77	[8.61, 8.93]	8.39	[7.93, 8.86]

Impulse response functions

- Want to compute $\dfrac{\partial \left(y_t - D(\theta) \right)}{\partial \varepsilon_1^k}(\theta)$

- Simply simulate the model!

  1. Set $\varepsilon_1^k = \sigma^k, \ \varepsilon_{k,t} = 0$for $t \geq 2$ and $\varepsilon_{j,t} = 0$, all $t \rightarrow$ Now we have constructed a sequence of $\varepsilon_t, \ t = 1, .., T$

  2. Use

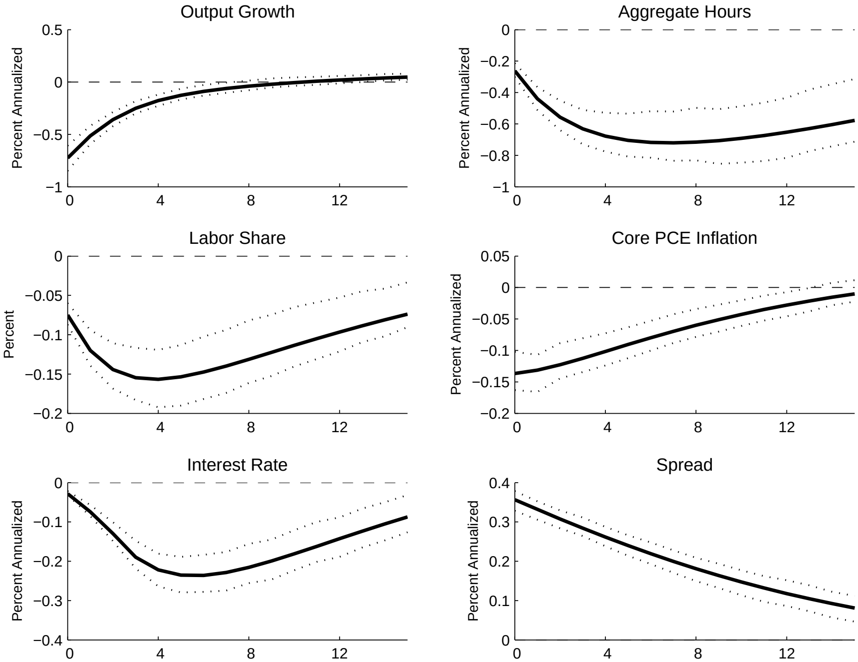
$$s_t = T(\theta)s_{t-1} + R(\theta)\varepsilon_t, \ t = 1, .., T$$

  to get the states, and

$$y_t - D(\theta) = Z(\theta)s_t + u_t, \ t = 1, .., T$$

  to get the y_t.

- Repeat for all draws $\theta^{(j)}, \ j = n^{burn} + 1, .., n^{sim}$.


Output Growth
Percent Annualized
Aggregate Hours
Percent Annualized
Labor Share
Percent
Core PCE Inflation
Percent Annualized
Interest Rate
Percent Annualized
Spread
Percent Annualized

Variance Decomposition

- Want to compute the fraction of the variance of y_t explained by shock $\varepsilon_{k,t}$
- Overall (unconditional) variance of y_t:

$$\mathrm{Var}(y_t) = Z(\theta)P_{0|0}Z(\theta)' + H(\theta)$$

where $P_{0|0}$ solves

$$P_{0|0} = T(\theta)P_{0|0}T(\theta)' + R(\theta)Q(\theta)R(\theta)'$$

- Variance of y_t attributed to shock k:

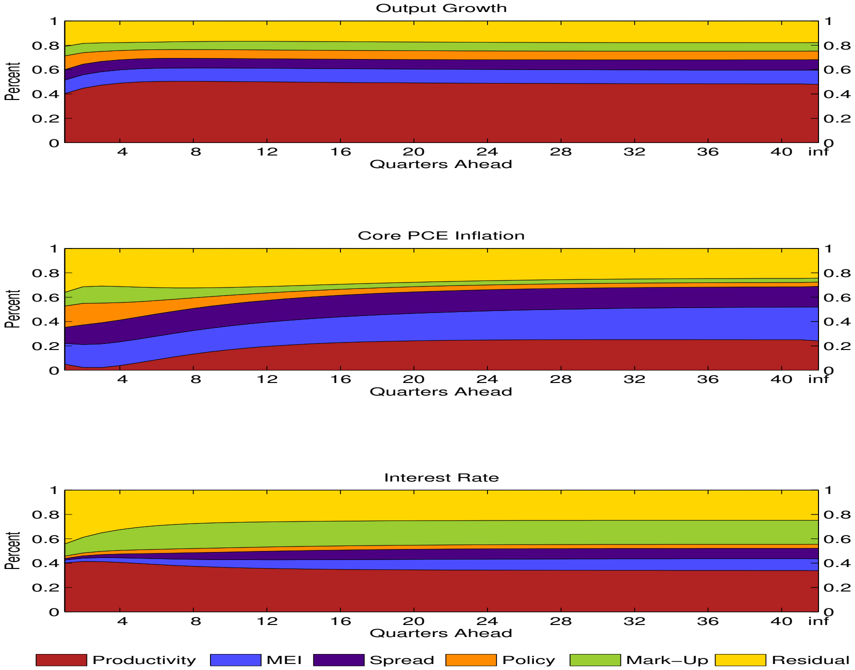
1. Construct $\tilde{Q}^k$ where all diagonal elements are set to 0 except for the k^{th}, which is equal to $\sigma^{k\ 2}$.

2. Compute the solution $P^k_{0|0}$ to

$$P^k_{0|0} = T(\theta)P^k_{0|0}T(\theta)' + R(\theta)\tilde{Q}^k R(\theta)'$$

and compute

$$\mathrm{Var}(y_t)^k = Z(\theta)P^k_{0|0}Z(\theta)'$$


Output Growth
Core PCE Inflation
Interest Rate
Percent
Quarters Ahead
0
0.2
0.4
0.6
0.8
1
4
8
12
16
20
24
28
32
36
40
inf
Productivity
MEI
Spread
Policy
Mark–Up
Residual

Inference on latent states

- What is the time series of the output gap, or r*? See Liberty St post on "Why Are Interest Rates So Low?" or see this presentation.

- Call $f(s_{0:T})$ any function mapping the vector of latent states into the object of interest, e.g. $r^*_{1:T} = \{Z_{r^f} s_1, .., Z_{r^f} s_T\}$, where Z_{r^f} selects the state corresponding to the real interest rate in the flexible price/wages economy. Then simply use the simulation smoother to obtain

$$p(f(s_{0:T})|y_{1:T}) = \int f(s_{0:T}) p(s_{0:T}|\theta, y_{1:T}) p(\theta|y_{1:T}) d(\theta, s_{0:T})$$

Shock decompositions

- What would history y_t have been, had only shock i hit the economy, and no other shock? See Liberty St post on "Developing a Narrative: The Great Recession and Its Aftermath"

1. Use the simulation smoother to compute draws $\varepsilon^{(j)}_{i,1:T}$, $j = n^{burn} + 1, .., n^{sim}$ from $p(\varepsilon_t | y_{1:T}, \theta)$, and $s_0^{(j)}$ from $p(s_0 | y_{1:T}, \theta)$.
2. Take the sequence of shock innovations for shock i, $\varepsilon^{(j)}_{i,1:T}$, and generate a new sequence of innovations $\tilde{\varepsilon}_{1:T}$ ($\tilde{\varepsilon}_t$ is of the same dimension as ε_t) by setting the i'th element $\tilde{\varepsilon}_{i,t} = \varepsilon^{(j)}_{i,t}$ for $t = 1, \ldots, T$
(and $\tilde{\varepsilon}_{i,t} \sim N(0, \sigma_i^2)$ for $t = T + 1, \ldots, T + H$, if interested in shock decomposition for forecasts).
All other elements of $\tilde{\varepsilon}_t$, $t = 1, \ldots, T + H$, are set equal to zero.

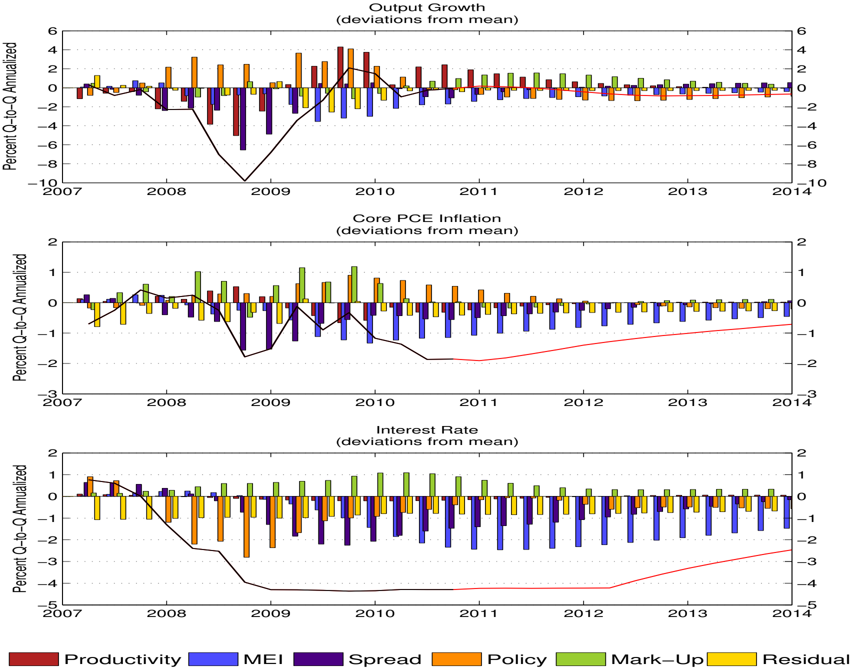
- If you are only interested in the mean shock decomposition you can use smoothed shocks $\varepsilon_{i,t|T}$ for each draw of θ.

3. Generate a counterfactual set of states $\tilde{s}_{1:T}$ from

$$\tilde{s}_t = T(\theta)\tilde{s}_{t-1} + R(\theta)\tilde{\varepsilon}_t, \;\; t = 1, .., T + H$$

and a counterfactual history $\tilde{y}_{1:T}$ from

$$\tilde{y}_t = D(\theta) + Z(\theta)\tilde{s}_t, \;\; t = 1, .., T + H.$$


Output Growth
(deviations from mean)
Percent Q-to-Q Annualized
6 4 2 0 −2 −4 −6 −8 −10
2007 2008 2009 2010 2011 2012 2013 2014
Core PCE Inflation
(deviations from mean)
Percent Q-to-Q Annualized
2 1 0 −1 −2 −3
2007 2008 2009 2010 2011 2012 2013 2014
Interest Rate
(deviations from mean)
Percent Q-to-Q Annualized
2 1 0 −1 −2 −3 −4 −5
2007 2008 2009 2010 2011 2012 2013 2014
Productivity
MEI
Spread
Policy
Mark−Up
Residual

Forecasting with DSGEs

- How do we generate forecasts $y_{T+1:T+H}$ from a state-space model? Same as any other state space model ...

$$p(y_{T+1:T+H}|y_{1:T}) =$$
$$\int_{(s_T,\theta)} p(y_{T+1:T+H}|s_T,\theta,y_{1:T}) \underbrace{p(s_T|\theta,y_{1:T})}_{\text{posterior of } s_T|\theta} \underbrace{p(\theta|y_{1:T})}_{\text{posterior of } \theta} d(s_T,\theta)$$

where

$$p(y_{T+1:T+H}|s_T,\theta,y_{1:T}) = \int_{s_{T+1:T+H}} p(y_{T+1:T+H}|s_{T+1:T+H}) \\ p(s_{T+1:T+H}|s_T,\theta,y_{1:T}) ds_{T+1:T+H}$$

In words...:

1. Use the Kalman filter to compute mean and variance of the distribution $p(s_T|\theta^{(j)}, y_{1:T})$. Generate a draw $s_T^{(j)}$ from this distribution, where $\theta^{(j)}$ is a draw from the posterior of θ.
2. Draw from $s_{T+1:T+H}|(s_T, \theta, y_{1:T})$ by generating a sequence of innovations $\epsilon_{T+1:T+H}^{(j)}$, and iterating the state transition equation forward starting from $s_T^{(j)}$:

$$s_t^{(j)} = T(\theta^{(j)})s_{t-1}^{(j)} + R(\theta^{(j)})\epsilon_t^{(j)}, \quad t = T+1, \dots, T+H.$$

3. Use the measurement equation to obtain $y_{T+1:T+H}^{(j)}$:

$$y_t^{(j)} = D(\theta^{(j)}) + Z(\theta^{(j)})s_t^{(j)}, \quad t = T+1, \dots, T+H. \quad \square$$

Why bother with forecasting with DSGE models?

- DSGE models have been trashed, bashed, and abused during the Great Recession and after. One of the many reasons for the bashing was their alleged inability to forecast.
- But DSGE models forecasts' accuracy is comparable to, if not better than, that of Blue Chip forecasters (and Greenbook)
- See Edge & Gürkaynak, BPEA 2010, and Del Negro & Schorfheide (2013 "DSGE Model-Based Forecasting," *Handbook of Economic Forecasting II*, also here)

Poverty of the econometrician's information set

- Quality of forecasts is constrained by quality of model, and the observables used by the econometrician. The "usual" set of observables (mostly NIPA based) falls short in two dimensions:

    1. Timeliness: NIPA data are available with a lag. Professional forecasters have current information that the DSGE econometrician is not using.

    2. Breadth: The "usual" set of observables may not convey enough information about the state of the economy.

- Augment the set of observables: Use nowcasts from professional forecasters, spreads, surveys ... $\rightarrow$ variables that may convey information about the state of the economy not contained in "usual" data set.

Real time data sets

- Level the playing field: don't give the DSGE econometrician information that private forecasters do not possess at the time of the forecasts (Croushore and Stark 2001, Edge and Gürkaynak 2010)

Quarter	Greenbook Date	End of Estimation Sample T	Initial Forecast Period $T+1$
Q1	Jan 21	2003:Q3 (F)	2003:Q4
	Mar 10	2003:Q4 (P)	2004:Q1
Q2	Apr 28	2003:Q4 (F)	2004:Q1
	June 23	2004:Q1 (P)	2004:Q2
Q3	Aug 4	2004:Q2 (A)	2004:Q3
	Sep 15	2004:Q2 (P)	2004:Q3
Q4	Nov 3	2004:Q3 (A)	2004:Q4
	Dec 8	2004:Q3 (P)	2004:Q4

Baseline DSGE Model: SW (2007)

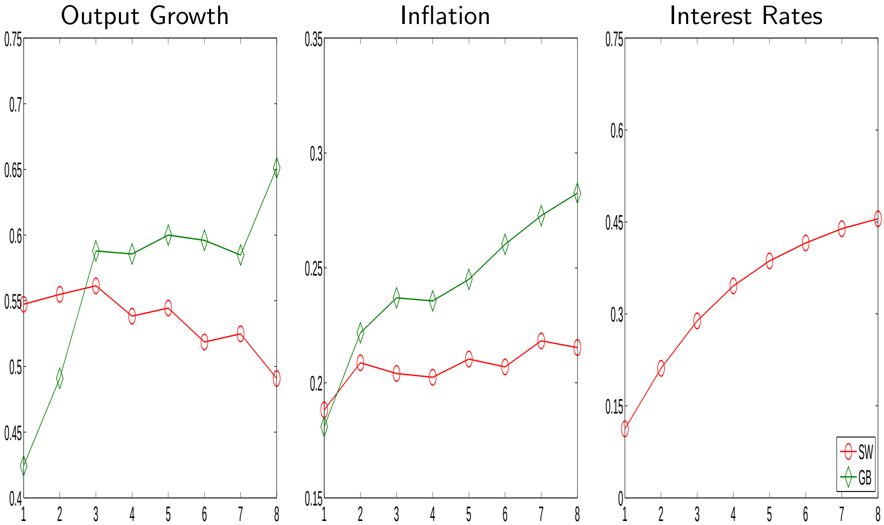
- Measurement equation:

$$
\begin{array}{lcl}
\textit{Output growth} & = & LN((GDPC)/LNSINDEX) * 100 \\
\textit{Consumption growth} & = & LN((PCEC/GDPDEF)/LNSINDEX) * 100 \\
\textit{Investment growth} & = & LN((FPI/GDPDEF)/LNSINDEX) * 100 \\
\textit{Real Wage growth} & = & LN(PRS85006103/GDPDEF) * 100 \\
\textit{Hours} & = & LN((PRS85006023 * CE16OV/100)/LNSINDEX) \\
 & & *100 \\
\textit{Inflation} & = & LN(GDPDEF/GDPDEF(-1)) * 100 \\
\textit{FFR} & = & FEDERAL\ FUNDS\ RATE/4
\end{array}
$$

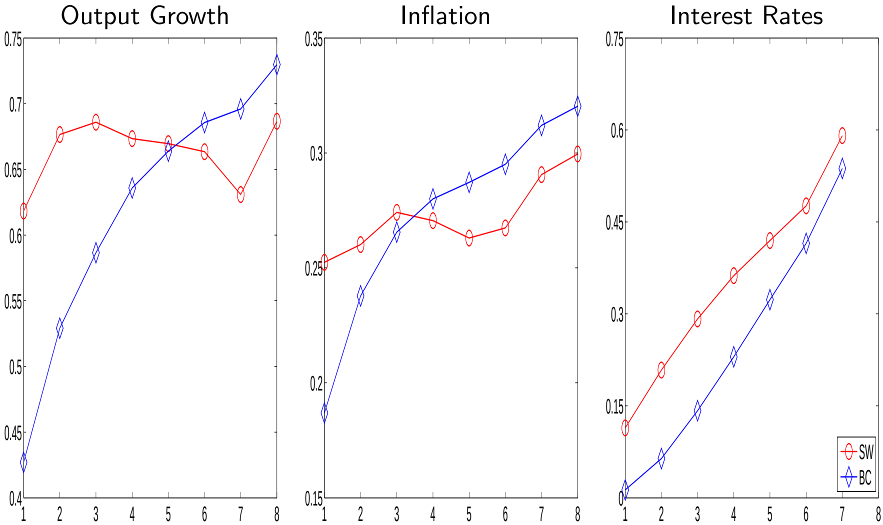
Sample starts in 1964:Q1

- Same prior on θ as SW.

SW vs Greenbook (March 1992-Sept 2004)



SW vs Blue Chip (Jan 1992-Apr 2011)



Incorporating 10-yrs inflation expectations from surveys

- SW forecasts inflation relatively well but ... somewhat tight prior on π^*:$\sim$ *Gamma*(.62, .10).
- No need of such a prior: Use a loose prior ($\pi^* \sim$ *Gamma*(.75, .40)) and survey data as an observable:

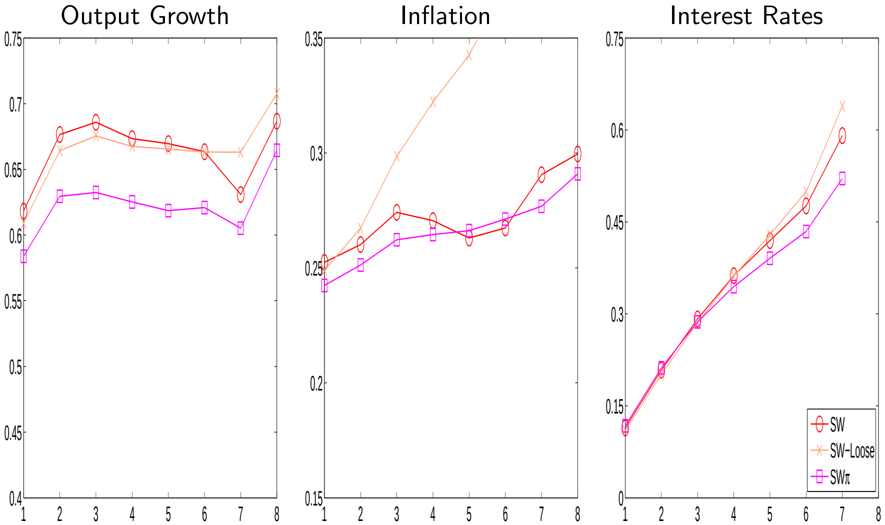
$$\pi_t^{O,40} \quad = \quad \pi_* + \boldsymbol{E}_t^{DSGE}\left[\frac{1}{40}\sum_{k=1}^{40}\pi_{t+k}\right]$$

- ... and change the model to be able to explain it:

$$\begin{aligned} R_t \quad &= \quad \rho_R R_{t-1} + (1-\rho_R)\left(\psi_1(\pi_t - \pi_t^*) + \psi_2(y_t - y_t^f)\right) \\ &\qquad +\psi_3\left((y_t - y_t^f) - (y_{t-1} - y_{t-1}^f)\right) + r_t^m, \end{aligned}$$

  where $\pi_t^* = \rho_{\pi^*}\pi_{t-1}^* + \sigma_{\pi^*}\epsilon_{\pi^*,t}$.
- Similar to Wright's "democratic prior" – but survey not used to form a prior.

SW vs SW-Loose vs SWπ

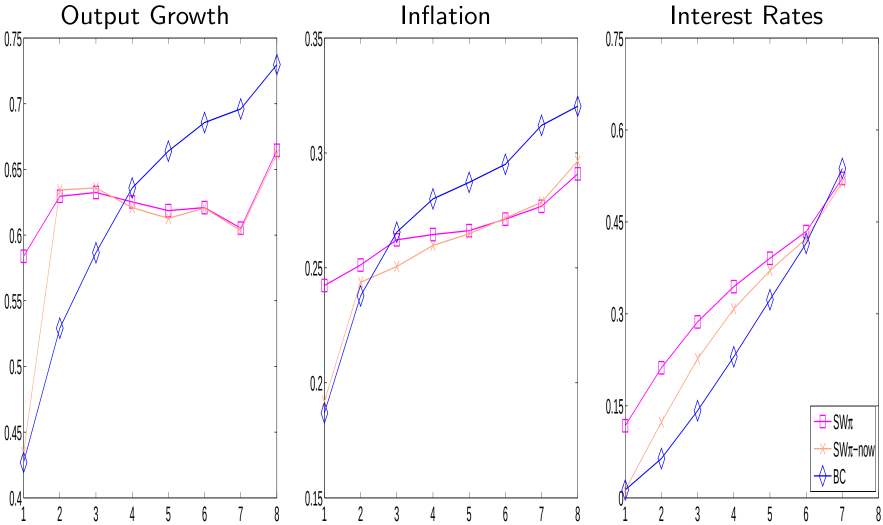


Timeliness of information: Incorporating nowcasts

- Factor model literature (for DSGEs, Boivin and Giannoni (2007)) addresses the issue by using the current indicators observed by professional forecasters (confidence indexes, ISM, durable goods orders, . . .) as data.

- As a shortcut, we use those data as digested by professional forecasters → incorporate Blue Chip consensus nowcasts as (possibly noisy) observations on GDP, inflation, ...

Incorporating nowcasts



Modeling Forward Guidance: Anticipated Policy Shocks

- We modify this rule to allow for forward guidance following Laseen & Svensson 2009:

$$\hat{R}_t = \rho_R \hat{R}_{t-1} + (1-\rho_R)\Big(\psi_\pi \sum_{j=0}^{3} \hat{\pi}_{t-j} + \psi_y \sum_{j=0}^{3} (\hat{y}_{t-j} - \hat{y}_{t-j-1} + \hat{z}_{t-j})\Big)$$
$$+ \epsilon_t^R + \sum_{k=1}^{K} \epsilon_{k,t-k}^R$$

  where $\epsilon_{k,t-k}^R$ is a policy shock that is known to agents at time $t-k$, but affects the policy rule k periods later, that is, at time t.

- Anticipated policy shocks are a simple way of capturing anticipated deviations from the standard reaction function

- Note: Even in the model, *not* commitment to a *path*: **conditionality** is still there!

Estimating Forward Guidance

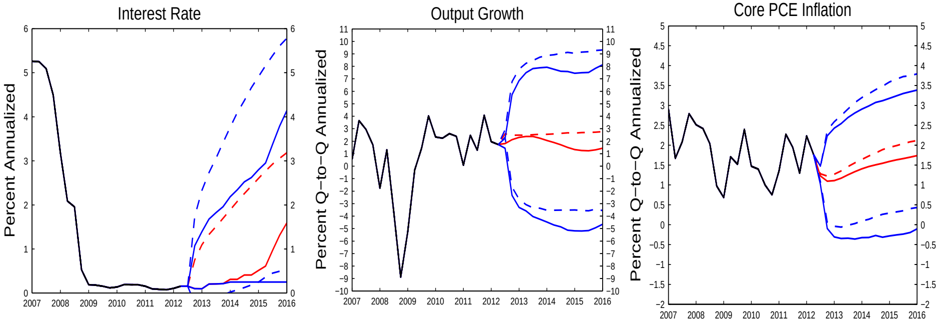
- Add Expected FFR to the measurement equations:

$$\begin{aligned} FFR^e_{t,t+k} &= 400\left(\boldsymbol{E}_t\widehat{R}_{t+k} + \ln R_*\right) \\ &= 400\left(Z_{R,.}(\theta)T(\theta)^k s_t + D_{R,.}(\theta)\right), \quad k = 1, .., K \end{aligned}$$

where $FFR^e_{t,t+k}$ is measured either using market expectations (e.g., OIS rates), or survey expectations (e.g., Blue Chip financial survey).

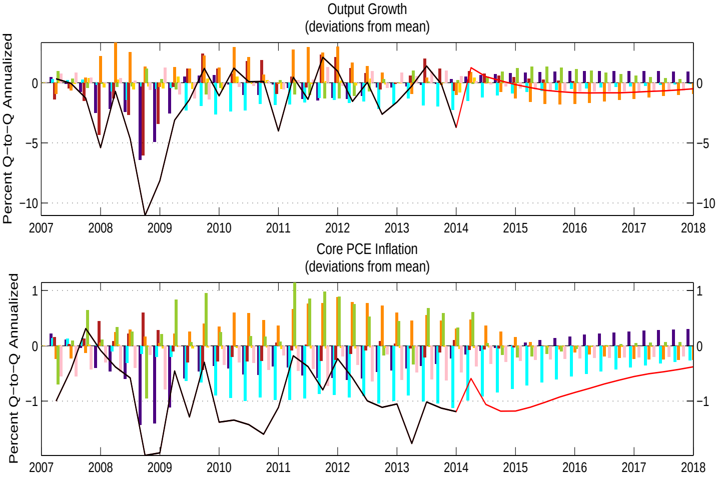
The Effect of Observing Expected Future Rates

- Introducing expected future rates in the measurement equation provides information to the econometrician on the state of the economy, which consists of both i) future policy shocks, ii) other latent variables → does not necessarily produce more optimistic forecasts



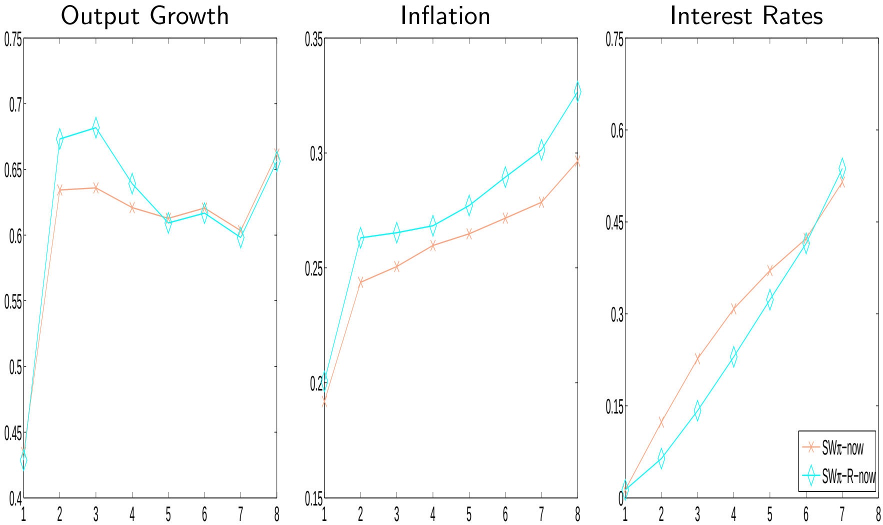
- Note: From the ex-post behavior of output and inflation the model should be able to tell whether the change in expected FFR is due to a policy shock or bad news

Historical Decomposition of Output and Inflation in the FRBNY DSGE Model



Black: data (2007Q1-2014Q4); red: forecast.

Forecasting using interest rate expectations



Implementing a Forward Guidance Policy Experiment

- Suppose that at the end of period T (after time T shocks are realized) the CB announces that, conditional on the state of the economy today $s_{T|T}$ (common knowledge), it expects the future path of interest-rates to be $\bar{R}_{T+1}, \ldots, \bar{R}_{T+\bar{H}}$.

- For the agents, the announcement is a one-time surprise in period $T+1$, corresponding to an unanticipated monetary policy shock ϵ^R_{T+1} and a sequence of anticipated shocks $\{\epsilon^R_{1,T+1}, \epsilon^R_{2,T+1}, \ \ldots, \ \epsilon^R_{K,T+1}\}$ where $K = \bar{H} - 1$.

- The solution to the following *linear* system of equations determines the time $T+1$ monetary policy shocks $\bar{\epsilon}^R = [\bar{\epsilon}^R_{T+1}, \bar{\epsilon}^{R'}_{1:K,T+1}]'$ as a function of the desired interest rate sequence $\bar{R}_{T+1}, \dots, \bar{R}_{T+\bar{H}}$

$$
\begin{aligned}
\bar{R}_{T+1} &= D_{R,.} + Z_{R,.} T s_{T|T} + Z_{R,.} R[\bar{\epsilon}^R_{T+1}, 0, \dots, 0, \bar{\epsilon}^{R\,'}_{1:K,T+1}]' \\
\bar{R}_{T+2} &= D_{R,.} + Z_{R,.} T^2 s_{T|T} + Z_{R,.} T R[\bar{\epsilon}^R_{T+1}, 0, \dots, 0, \bar{\epsilon}^{R\,'}_{1:K,T+1}]' \\
&\vdots \\
\bar{R}_{T+\bar{H}} &= D_{R,.} + Z_{R,.} T^{\bar{H}} s_{T|T} + Z_{R,.} (T)^{\bar{H}-1} R[\bar{\epsilon}^R_{T+1}, 0, \dots, 0, \bar{\epsilon}^{R\,'}_{1:K,T+1}]'
\end{aligned}
$$

- Iterate forward the state transition equation starting from $s_{T|T}$ plugging in the policy shocks $\bar{\epsilon}^R$ in period $T+1$

$$s_{T+1|T} = T(\theta^{(j)})s_{T+1|T} + R(\theta^{(j)})[\epsilon_t^R, 0, \ldots, 0, \epsilon_{1:K,t}^{R\;\prime}]'$$

and no shocks afterwards

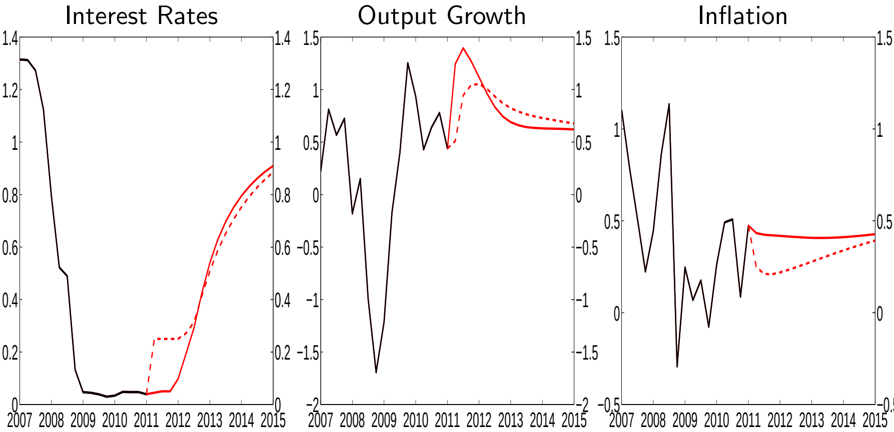
$$s_{t|T} = T(\theta^{(j)})s_{t-1|T}, \quad t = T+2, \ldots, T+H,$$

(note, the transition equation will take care of putting the anticipated shocks into the future policy rule)

- and plug the future states into the measurement equation to get the impact on output, inflation ...

- See section 6.3 in Del Negro, Schorfheide ("DSGE Model Forecasting", Handbook of Forecasting)

... forecasts conditional on an FFR path



Forecasting the Great Recession

- In addition to the SW model, we now consider a model with financial frictions along the lines of Bernanke, Gertler, Gilchrist (1999).
- Gross nominal return on capital:

$$\tilde{R}_t^k = \lambda r_t^k + (1-\lambda) q_t^k - q_{t-1}^k + \pi_t$$

- SW model: arbitrage condition between return on capital and return on nominal bond:

$$\mathbb{E}_t[\tilde{R}_{t+1}^k] = R_t + b_t,$$

where $\tilde{R}_t^k$ is treated as latent and b_t is a shock.
- SW-FF Model: arbitrage condition is

$$\mathbb{E}_t[\tilde{R}_{t+1}^k] = R_t + b_t + \zeta_{sp,b}\left(q_t^k + \bar{k}_t - n_t\right) + \tilde{\sigma}_{\omega,t}$$

where $\tilde{R}_t^k - R_t$ is treated as observed, $\tilde{\sigma}_{\omega,t}$ is an additional shock, and n_t is an additional endogenous variable.

Forecasting the Crisis: Model Versions

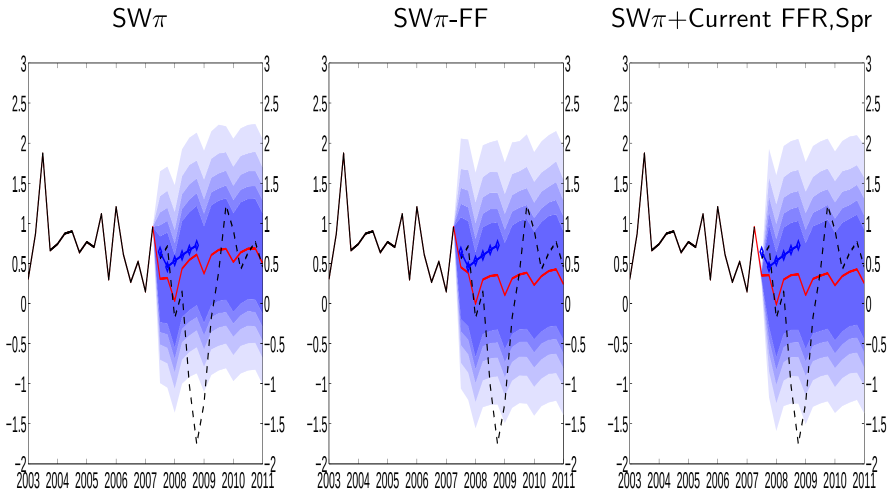
- SWπ: Smets-Wouters model with time-varying inflation target anchored by long-run inflation expectations. We do NOT use external nowcasts here.

- SWπ-FF: Smets-Wouters model with time-varying inflation target anchored by long-run inflation expectations and financial frictions. Utilizes data on spreads until period T.

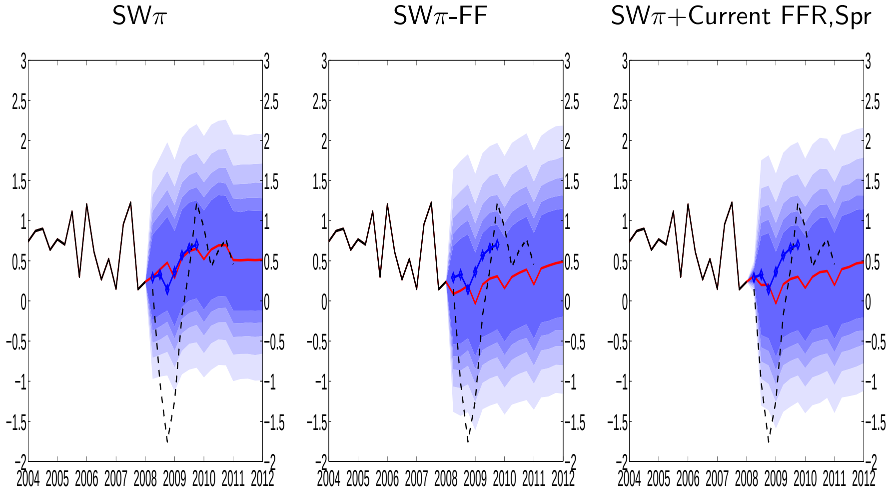
- SWπ-FF-Current: Smets-Wouters model with time-varying inflation target anchored by long-run inflation expectations and financial frictions. Also use FFR and spread from current quarter $T+1$.

- Spreads: based on Baa bonds versus 10-year treasury rate.

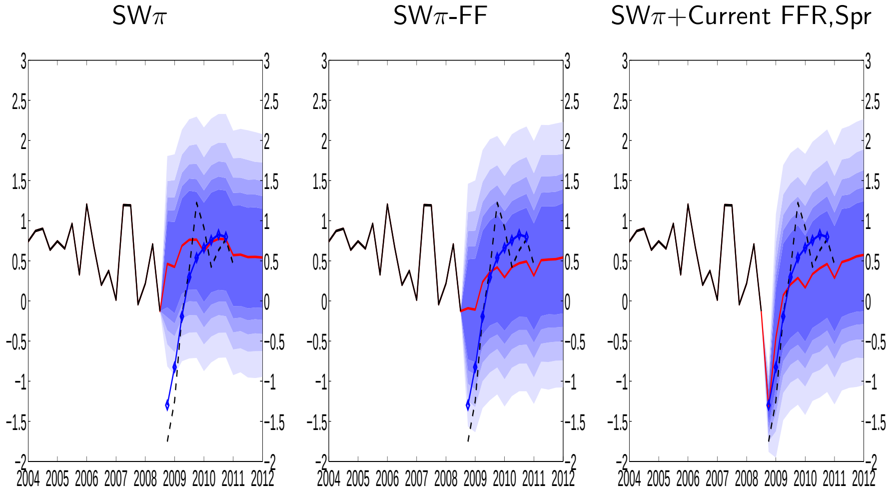
Forecasting the Great Recession: Oct 10, 2007 (2007Q2 data)



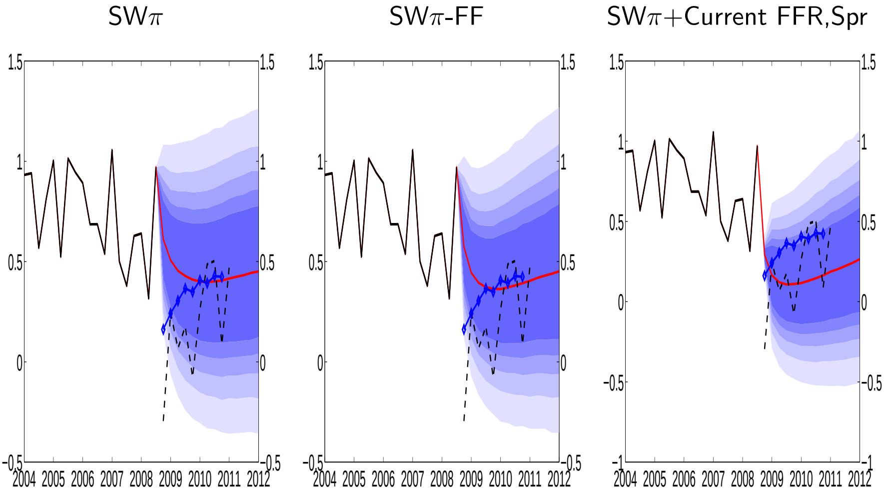
July 10, 2008 (2008Q1 data)



Jan 10, 2009 (2008Q3 data)



Forecasting the Great Recession: Inflation



See Del Negro, Giannoni, Schorfheide, Inflation in the Great Recession and New Keynesian Models, AEJ Macro 2015

Evaluation

- Question: are predictive densities are well calibrated?
- Roughly: in a sequential forecasting setting events that are predicted to have 20% probability, should roughly occur 20% of the time.
- Probability Integral Transforms:
  - If Y is cdf $F(y)$, then

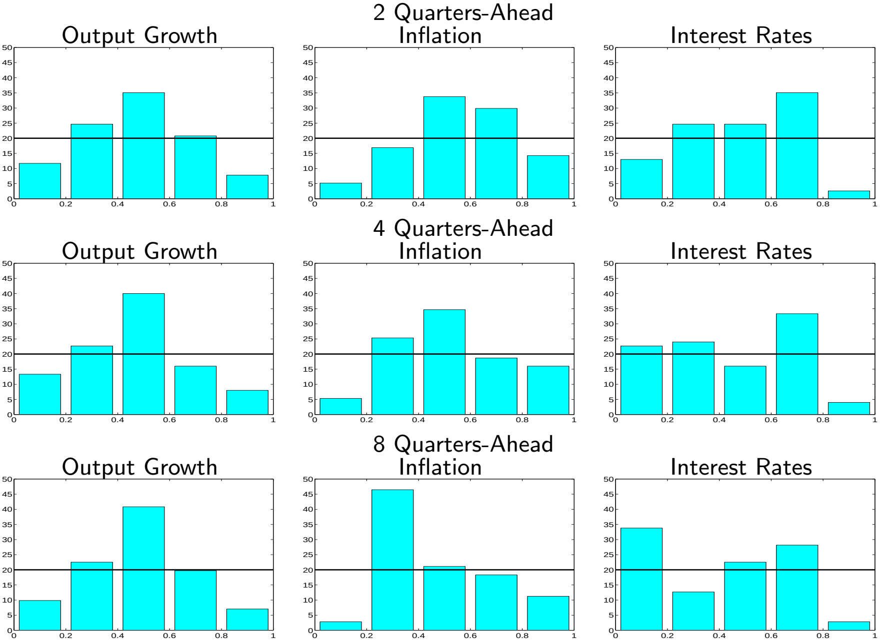
$$\mathbb{P}\{F(Y) \leq z\} = \mathbb{P}\{Y \leq F^{-1}(z)\} = F\big(F^{-1}(z)\big) = z$$

  - PITs

$$z_{i,t,h} = \int_{-\infty}^{y_{i,t+h}} p(\tilde{y}_{i,t+h}|Y_{1:T}) d\tilde{y}_{i,t+h}.$$

References for PITs: Rosenblatt (1952), Dawid (1984), Kling and Bessler (1989), Diebold, Gunther, and Tay (1998), Diebold, Hahn, and Tay (1999), ..., Geweke and Amisano (2010), Herbst and Schorfheide (2011).

PITs


2 Quarters-Ahead
Output Growth
Inflation
Interest Rates
4 Quarters-Ahead
Output Growth
Inflation
Interest Rates
8 Quarters-Ahead
Output Growth
Inflation
Interest Rates

Model Comparison

- Question: Does model $\mathcal{M}_1$ fit better than model $\mathcal{M}_2$?

- In a Bayesian framework, model comparison is conducted using Posterior Odds:

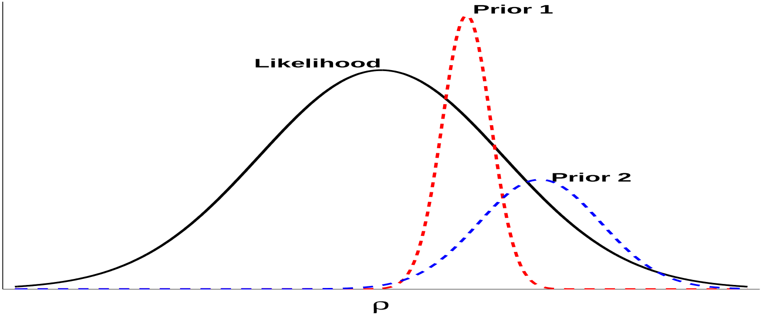
$$\underbrace{\frac{p(\mathcal{M}_1|y_{1:T})}{p(\mathcal{M}_2|y_{1:T})}}_{\text{Posterior Odds}} = \underbrace{\frac{p(y_{1:T}|\mathcal{M}_1)}{p(y_{1:T}|\mathcal{M}_2)}}_{\text{Bayes Factor}} \underbrace{\frac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)}}_{\text{Prior Odds}}$$

- Bayes Factor – the ratio of marginal likelihoods – summarizes the sample information as to which model achieves the best fit.

Priors and Bayesian Model Comparisons

- The marginal likelihood (or marginal data density) is the likelihood of observing the data under model $\mathcal{M}_i$, and is computed as the integral of the likelihood with respect to the prior:

$$p(y_{1:T}|\mathcal{M}_i) = \int \underbrace{p(y_{1:T}|\,\theta, \mathcal{M}_i)}_{\text{Likelihood}} \underbrace{p(\theta, \mathcal{M}_i)}_{\text{Prior}} d\theta$$



- Lindley's paradox: flat (or almost flat) priors can kill any model, no matter how well it fits the data.

Computing the marginal likelihood

- Geweke's modified harmonic mean estimator
- Harmonic mean estimators are based on the following identity

$$\frac{1}{p(y_{1:T})} = \int \frac{f(\theta)}{p(y_{1:T}|\theta)p(\theta)} p(\theta|y_{1:T})d\theta,$$

where $\int f(\theta)d\theta = 1$.

- Conditional on the choice of $f(\theta)$ an obvious estimator is

$$\hat{p}_G(y_{1:T}) = \left[\frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} \frac{f(\theta^{(j)})}{p(y_{1:T}|\theta^{(j)})p(\theta^{(j)})} \right]^{-1}$$

where $\theta^{(j)}$ is drawn from the posterior $p(\theta|y_{1:T})$.

- Geweke (1999):

$$\begin{aligned} f(\theta) &= \tau^{-1}(2\pi)^{-d/2}|V_\theta|^{-1/2} \exp\left[-0.5(\theta-\bar{\theta})' V_\theta^{-1}(\theta-\bar{\theta})\right] \\ &\quad \times \left\{(\theta-\bar{\theta})' V_\theta^{-1}(\theta-\bar{\theta}) \leq F^{-1}_{\chi^2_d}(\tau)\right\}. \end{aligned}$$