Dynamic Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance

Marco Del Negro, Raiden B. Hasegawa

Federal Reserve Bank of New York

Frank Schorfheide

University of Pennsylvania

8th ECB Workshop on Forecasting Techniques

Motivation

- Model uncertainty is pervasive.

- How should we combine models for forecasting, policy analysis, risk assessment – especially when we suspect that their relative appeal may change over time (e.g., financial crisis vs ‘normal’ times)?

Application

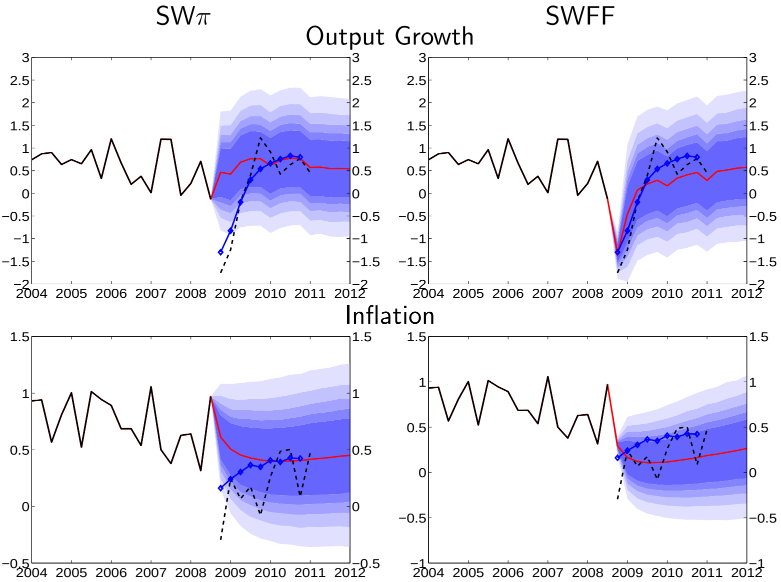
- Are financial variables (spreads) useful in forecasting macro economic outcomes?

  - Many macroeconomists paid scant attention to financial frictions models before the recent crisis. Why? Did these models not forecast very well in ‘normal’ times?

  - Which models should policymakers use now?

- We focus in particular on forecasting with DSGE models (Del Negro Schorfheide 2013, Handbook of Economic Forecasting).

DSGE forecasts of the Great Recession



Methodological Contribution

- Geweke and Amisano 2011, Hall and Mitchell 2007, *optimal pools*:

$$\max_{\lambda \in [0,1]} \prod_{t=1}^{T} \left(\lambda p(y_t | \mathcal{I}_{t-1}, \mathcal{M}_1) + (1-\lambda) p(y_t | \mathcal{I}_{t-1}, \mathcal{M}_2) \right)$$

static
$\Downarrow$
dynamic

$$p(y_t | \mathcal{I}_{t-1}, \lambda_t) = \lambda_t p(y_t | \mathcal{I}_{t-1}, \mathcal{M}_1) + (1-\lambda_t) p(y_t | \mathcal{I}_{t-1}, \mathcal{M}_2), \quad \lambda_t \in [0,1]$$

  - Related approaches: Billio et al. 2013, Raftery et al. 2013.

- Conduct inference on the time series of λ_t and ask: Is there time variation? Was the weight skewed toward one model vs the other before the Great Recession? What is λ_t now?

Combining Models - A Stylized Framework

- We have a principal-agent setting in mind...
- Agents = econometric modelers = Coenen, Jarocinski, Lenza... who provide principal with predictive densities $p(y_t|y_{1:t-1}, \mathcal{M}_i)$.
- Principal = policy maker = Smets, ... who aggregates information obtained from modelers.
- Agents are rewarded based on the realized value of $\ln p(y_t|y_{1:t-1}, \mathcal{M}_i)$ (induces truth-telling).

Bayesian Model Averaging (BMA)

- At any time T the policy maker can use the predictive densities to form marginal likelihoods:

$$p(y_{1:T}|\mathcal{M}_i) = \prod_{t=1}^{T} p(y_t|y_{1:t-1}, \mathcal{M}_i)$$

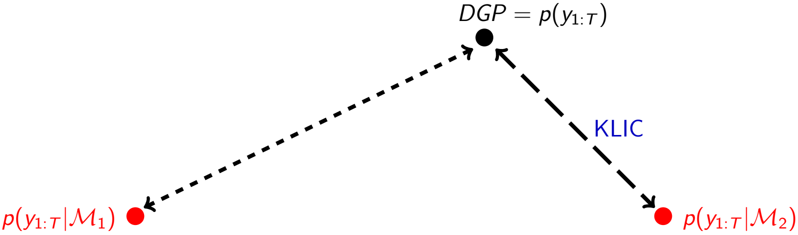
- ...use them to update model probabilities:

$$\lambda_T^{BMA} = \frac{\lambda_0^{BMA} p(y_{1:T}|\mathcal{M}_1)}{\lambda_0^{BMA} p(y_{1:T}|\mathcal{M}_1) + (1-\lambda_0^{BMA}) p(y_{1:T}|\mathcal{M}_2)}$$

where $\lambda_T^{BMA} = \mathbb{P}[\mathcal{M}_1 \text{ is correct}|y_{1:T}]$ and let λ_0^{BMA} be prior probability of $\mathcal{M}_1$.

BMA and Model Misspecification

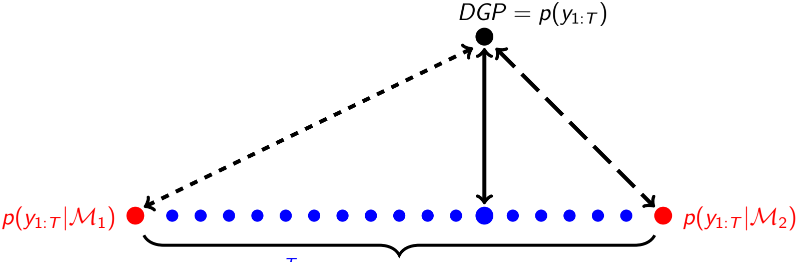
- This approach is based on the assumption that the model space contains the 'true' model ("complete model space")



- $\lambda_T^{BMA} \xrightarrow{a.s.} 1$ or 0 as $T \longrightarrow \infty$ (Dawid 1984): Asymptotically, no model averaging!
- If the policy maker mistrusts all of the models in the pool, this may not be the best approach.

Optimal Pools vs BMA

- A policy maker concerned about misspecification of $\mathcal{M}_i$ could create convex combinations of predictive densities:

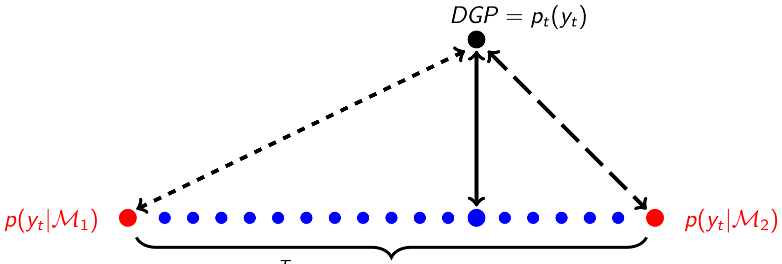


$$p(y_{1:T}|\lambda, \mathcal{M}_\lambda) = \prod_{t=1}^{T} \left\{ \lambda \, p\left(y_t|y_{1:t-1}, \mathcal{M}_1\right) + (1-\lambda) \, p\left(y_t|y_{1:t-1}, \mathcal{M}_2\right) \right\}$$

- $\lambda_T^{SP} = \text{argmax}_{\lambda \in [0,1]} p(y_{1:T}|\lambda, \mathcal{M}_\lambda)$ generally $\not\to$ 1 or 0 (unless one of the models is correct): Exploits gains from diversification.
- In a time-varying setting, policy maker needs to make inference with respect to $\lambda_{1:T}$.

Optimal Pools vs BMA

- A policy maker concerned about misspecification of $\mathcal{M}_i$ could create convex combinations of predictive densities:

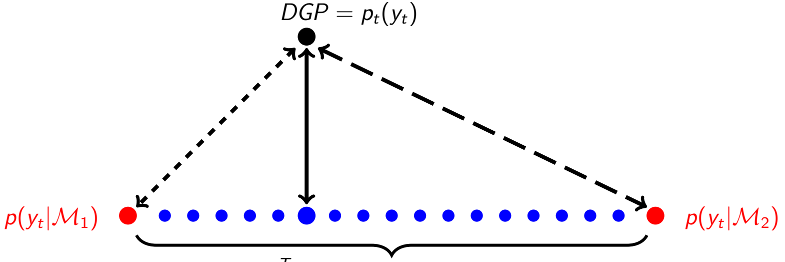


$$p(y_{1:T}|\lambda_{1:T},\mathcal{M}_\lambda) = \prod_{t=1}^{T}\left\{\lambda_t\ p\left(y_t|y_{1:t-1},\mathcal{M}_1\right) + (1-\lambda_t)\ p\left(y_t|y_{1:t-1},\mathcal{M}_2\right)\right\}$$

- $\lambda_T^{SP} = \text{argmax}_{\lambda\in[0,1]} p(y_{1:T}|\lambda,\mathcal{M}_\lambda)$ generally $\not\to$ 1 or 0 (unless one of the models is correct): Exploits gains from diversification.
- In a time-varying setting, policy maker needs to make inference with respect to $\lambda_{1:T}$.

Optimal Pools vs BMA

- A policy maker concerned about misspecification of $\mathcal{M}_i$ could create convex combinations of predictive densities:



$$p(y_{1:T}|\lambda_{1:T}, \mathcal{M}_\lambda) = \prod_{t=1}^{T} \left\{ \lambda_t \; p\left(y_t|y_{1:t-1}, \mathcal{M}_1\right) + (1-\lambda_t) \; p\left(y_t|y_{1:t-1}, \mathcal{M}_2\right) \right\}$$

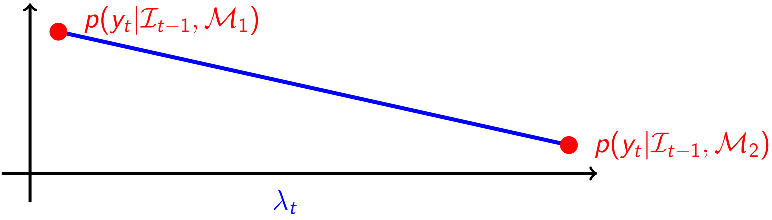
- $\lambda_T^{SP} = \text{argmax}_{\lambda \in [0,1]} p(y_{1:T}|\lambda, \mathcal{M}_\lambda)$ generally $\not\to$ 1 or 0 (unless one of the models is correct): Exploits gains from diversification.
- In a time-varying setting, policy maker needs to make inference with respect to $\lambda_{1:T}$.

Dynamic Pools – Likelihood Function

- Likelihood function:

$$\prod_{t=1}^{T} p(y_t|\mathcal{I}_{t-1}, \lambda_{1:T}) = \prod_{t=1}^{T} \Big[\lambda_t p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_1) + (1-\lambda_t) p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_2) \Big].$$

- Period t contribution to likelihood looks like:



- Need (stochastic process) prior for sequence $\lambda_{1:T}$.

Dynamic Pools - (Hierarchical) Prior

- Prior $p(\lambda_{1:T}|\rho)$ for sequence $\lambda_{1:T}$:

$$\begin{aligned} x_t &= \rho x_{t-1} + \sqrt{1-\rho^2}\varepsilon_t, \quad \varepsilon_t \sim iid\ N(0,1), \quad x_0 \sim N(0,1), \\ \lambda_t &= \Phi(x_t) \end{aligned}$$

  where $\Phi(.)$ is the Gaussian CDF.

- ρ controls the amount of "smoothing."

- As $\rho \longrightarrow 1$: dynamic pool $\longrightarrow$ static pool.

- Unconditionally, $\lambda_t \sim U[0,1]$ for all t.

Dynamic Pools - Nonlinear State Space System

- Measurement equation:

$$p(y_t|\mathcal{I}_{t-1}, \lambda_t) = \lambda_t p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_1) + (1-\lambda_t)p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_2)$$

- Transition equation:

$$\lambda_t = \Phi(x_t), \quad x_t = \rho x_{t-1} + \sqrt{1-\rho^2}\varepsilon_t, \quad \varepsilon_t \sim iid\ N(0,1)$$

- Use particle filter to construct the sequence $p(\lambda_t|\rho, y_{1:t})$.

- Predictive density:

$$\begin{aligned} & p(y_t|y_{1:t-1}) \\ &= \int \Big\{ \lambda_t p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_1) \\ &\quad +(1-\lambda_t)p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_2) \Big\} p(\lambda_t|y_{1:t-1}, \rho)p(\rho|y_{1:t-1})d\lambda_t d\rho \\ &= \hat{\lambda}_{t|t-1}p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_1) + (1-\hat{\lambda}_{t|t-1})p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_2), \end{aligned}$$

where $\hat{\lambda}_{t|t-1} = \mathbb{E}[\lambda_t|y_{1:t-1}]$.

Dynamic Pools - Nonlinear State Space System

- Measurement equation:

$$p(y_t|\mathcal{I}_{t-1}, \lambda_t) = \lambda_t p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_1) + (1-\lambda_t)p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_2)$$

- Transition equation:

$$\lambda_t = \Phi(x_t), \quad x_t = \rho x_{t-1} + \sqrt{1-\rho^2}\varepsilon_t, \quad \varepsilon_t \sim iid\ N(0,1)$$

- Use particle filter to construct the sequence $p(\lambda_t|\rho, y_{1:t})$.
- Predictive density:

$$\begin{aligned} & p(y_t|y_{1:t-1}) \\ &= \int \Big\{ \lambda_t p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_1) \\ &\quad + (1-\lambda_t)p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_2) \Big\} p(\lambda_t|y_{1:t-1}, \rho) p(\rho|y_{1:t-1}) d\lambda_t d\rho \\ &= \hat{\lambda}_{t|t-1} p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_1) + (1-\hat{\lambda}_{t|t-1}) p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_2), \end{aligned}$$

where $\hat{\lambda}_{t|t-1} = \mathbb{E}[\lambda_t|y_{1:t-1}]$.

Alternative Law of Motions for the Weights (λ_t)

- Allow for a mean μ in

$$\begin{aligned} x_t &= \mu + \rho x_{t-1} + \sqrt{1-\rho^2}\varepsilon_t, \ \varepsilon_t \sim iid\ N(0,1), \ x_0 \sim N(0,1), \\ \lambda_t &= \Phi(x_t) \end{aligned}$$

  so that $\rho < 1 \not\to \boldsymbol{E}[\lambda_t] = 1/2$ for $t \to \infty$.
  - $\mu \sim \mathcal{N}(0, \sigma_\mu^2)$ being a natural prior

- ... and/or a standard deviation $\sigma \neq 1$

$$x_t = \rho x_{t-1} + \sigma \sqrt{1-\rho^2}\varepsilon_t,$$

- $\sigma = 1$: flat
- $\sigma < 1$: $\bigcap$ shaped (pull toward equal weights).
- $\sigma > 1$: $\bigcup$ shaped (pull away from equal weights).

Alternative Law of Motions for the Weights (λ_t)

- Allow for a mean μ in

$$\begin{aligned} x_t &= \mu + \rho x_{t-1} + \sqrt{1-\rho^2}\varepsilon_t, \; \varepsilon_t \sim iid\ N(0,1), \; x_0 \sim N(0,1), \\ \lambda_t &= \Phi(x_t) \end{aligned}$$

  so that $\rho < 1 \not\to \boldsymbol{E}[\lambda_t] = 1/2$ for $t \to \infty$.
  - $\mu \sim \mathcal{N}(0, \sigma_\mu^2)$ being a natural prior

- ... and/or a standard deviation $\sigma \neq 1$

$$x_t = \rho x_{t-1} + \sigma \sqrt{1-\rho^2}\varepsilon_t,$$

- $\sigma = 1$: flat
- $\sigma < 1$: $\bigcap$ shaped (pull toward equal weights).
- $\sigma > 1$: $\bigcup$ shaped (pull away from equal weights).

- Markov-Switching setup

$$p(y_t|\mathcal{I}_{t-1},\lambda_t) = s_t p(y_t|\mathcal{I}_{t-1},\mathcal{M}_1) + (1-s_t)p(y_t|\mathcal{I}_{t-1},\mathcal{M}_2),$$
$$s_t = 0,1,\ \pi_{ij} = P(s_t = i|s_{t-1} = j)$$

so that λ_t could be interpreted as the probability of being in regime 1 (i.e., $\lambda_t = P(s_t|y_{1:t})$ for filtered).

Reestimate Models for Each $\lambda_{1:T}$?

- Full information Bayes' answer: Yes (e.g., Waggoner and Zha, 2012)

- Joint distribution:

$$p(y_{1:T}, \lambda_{1:T}, \theta_{(1)}, \theta_{(2)}) = \left(\prod_{t=1}^{T} \lambda_t p(y_t | y_{1:t-1}, \theta_{(1)}, \mathcal{M}_1) + (1 - \lambda_t) p(y_t | y_{1:t-1}, \theta_{(2)}, \mathcal{M}_2) \right) \times p(\theta_{(1)}) p(\theta_{(2)}) p(\lambda_{1:T})$$

$\rightarrow$ posterior of $(\theta_{(1)}, \theta_{(2)})$ will generally depend on $\lambda_{1:T}$:

$$p(\theta_{(1)}, \theta_{(2)}, \lambda_{1:T} | y_{1:T}) = p(\theta_{(1)}, \theta_{(2)} | y_{1:T}, \lambda_{1:T}) p(\lambda_{1:T} | y_{1:T}).$$

Our approach: No

1. Models are in general not estimated on the same variables (e.g., DSGEs vs VARs vs FRBUS) $\rightarrow$ Likelihood

$$p(y_t|y_{1:t-1}, \theta_{(m)}, \mathcal{M}_m)$$

not available for the same set of variables for all models.

2. Principal may only care about a subset of these variables in drawing inference about $\lambda_{1:T}$
   - Let $\mathcal{I}_t = \{y_{1:t}, z_{1:t}\}$.
   - For any set of beliefs $p(z_t|y_t, \mathcal{I}_{t-1}, \mathcal{P})$, $t = 1, \ldots, T$, the policy maker can form the joint distribution:

$$p(y_{1:T}, z_{1:T}|\lambda_{1:T}) = \prod_{t=1}^{T} \Big\{ \big(\lambda_t p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_1) + (1-\lambda_t) p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_2)\big) \times p(z_t|y_t, \mathcal{I}_{t-1}, \mathcal{P}) \Big\}$$

which implies

$$p(\lambda_{1:T}|\mathcal{I}_T) \propto p(\lambda_{1:T}) \prod_{t=1}^{T} \big(\lambda_t p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_1) + (1-\lambda_t) p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_2)\big).$$

Our approach: No

1. Models are in general not estimated on the same variables (e.g., DSGEs vs VARs vs FRBUS) $\rightarrow$ Likelihood

$$p(y_t|y_{1:t-1}, \theta_{(m)}, \mathcal{M}_m)$$

not available for the same set of variables for all models.

2. Principal may only care about a subset of these variables in drawing inference about $\lambda_{1:T}$
   - Let $\mathcal{I}_t = \{y_{1:t}, z_{1:t}\}$.
   - For any set of beliefs $p(z_t|y_t, \mathcal{I}_{t-1}, \mathcal{P})$, $t = 1, \ldots, T$, the policy maker can form the joint distribution:

$$p(y_{1:T}, z_{1:T}|\lambda_{1:T}) = \prod_{t=1}^{T} \Big\{ \big(\lambda_t p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_1) + (1-\lambda_t) p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_2)\big) \times p(z_t|y_t, \mathcal{I}_{t-1}, \mathcal{P}) \Big\}$$

   which implies

$$p(\lambda_{1:T}|\mathcal{I}_T) \propto p(\lambda_{1:T}) \prod_{t=1}^{T} \big(\lambda_t p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_1) + (1-\lambda_t) p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_2)\big).$$

3. It is not always possible to re-estimate all models for any sequence $\lambda_{1:T}$

   - Estimation of large-scale DSGE models is computationally costly, and for some models that are not estimated using likelihood based methods (e.g., FRBUS) it is not clear how they would incorporate the information coming from $\lambda_{1:T}$.
   - Agents/modelers provide principal with predictive densities $p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_m)$. The agents are rewarded based on the realized value of $\ln p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_m)$
   - Principal aggregates (limited) information obtained from modelers.

Multi-Step Forecasting

- Policy maker may be interested in multi-step forecasts (and so are we): e.g., average growth/inflation over h periods:

$$\bar{y}_{t,h} = \frac{1}{h}\sum_{s=0}^{h-1} y_{t-s}.$$

- From a full information Bayesian perspective, the policy maker should construct the posterior of $\lambda_{1:T}$ based on $h = 1$ (one-step-ahead predictive densities): $p(y_t|\mathcal{I}_{t-1}, \mathcal{M}_m)$.
- However, if model misspecification is a serious concern, then it is reasonable to use the loss function, i.e., the h-step-ahead predictive densities to determine $\lambda_{1:T}$.
    - Literature on loss-function-based versus likelihood-based estimation of forecasting model, e.g., Schorfheide (JoE, 2005).

- As is common in the literature on predictive regressions $y_t = \beta_0 + \beta_1 x_{t-h} + u_t$, we estimate the pooling weights directly, ignoring the overlap between $\bar{y}_{t,h}$, $\bar{y}_{t-1,h}$, etc.

- (Pseudo) likelihood:

$$p^{(h)}(\bar{y}_{1:T,h}|\lambda_{1:T}) = \prod_{t=1}^{T} \left[\lambda_t^{(h)} p(\bar{y}_{t,h}|\mathcal{I}_{t-h}, \mathcal{M}_1) + (1-\lambda_t^{(h)}) p(\bar{y}_{t,h}|\mathcal{I}_{t-h}, \mathcal{M}_2) \right]$$

- Particle filter generates pseudo posteriors $p^{(h)}(\lambda_t^{(h)}|\rho, \bar{y}_{1:t,h})$.

- We use the following predictive density for forecasting (with $h = 4$):

$$\begin{aligned} p^{(h)}(\bar{y}_{t,h}|\rho, \bar{y}_{1:t-h,h}) &= \hat{\lambda}_{t|t-h}^{(h)} p(\bar{y}_{t,h}|\mathcal{I}_{t-h}, \mathcal{M}_1) \\ &\quad + (1 - \hat{\lambda}_{t|t-h}^{(h)}) p(\bar{y}_{t,h}|\mathcal{I}_{t-h}, \mathcal{M}_2). \end{aligned}$$

The Data

- Forecast for each model are based on real-time data (sample starts in 1964:Q1)
- Information sets ($\mathcal{I}_t^m$) for forecasts:
  - SWπ: output growth, inflation, fed funds, consumption growth, investment growth, wage growth, hours worked, 10-yrs Inflation Expectations from Surveys.
  - SWπFF: $\ldots +$ spread
  - $\mathcal{I}_t^m$ includes current $(t+1)$ values for financial variables.
- Variables to be forecast (y_t): output growth, inflation
- Forecast evaluation period $(t = 1, .., T)$: 1992:Q1-2011:Q2

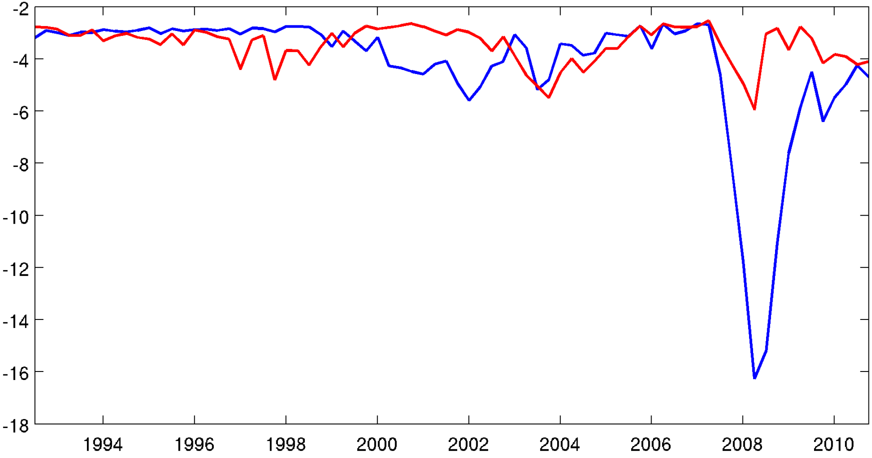
Questions

1. Inference on λ_t: Is there significant time variation in the relative forecasting performance of the two models, as captured by the estimated distribution of λ_t?

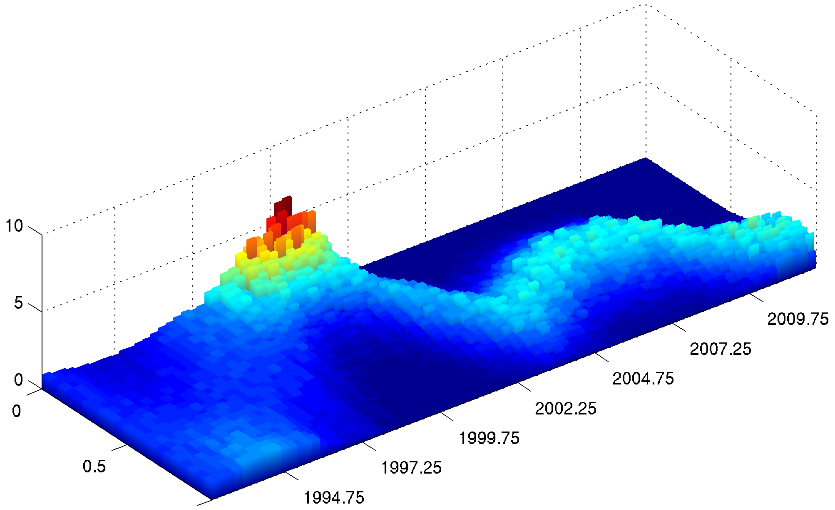
   - Does λ_t change rapidly enough when estimated in real time to offer useful guidance to policy makers or forecasters?

2. Forecasting performance: Do the dynamic pools perform better in real time than forecasting with static pools, BMA weights, equal weights?

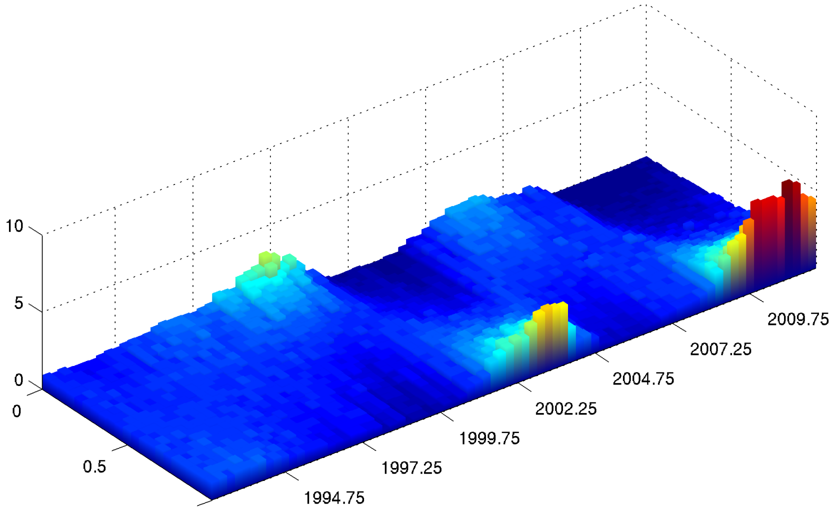
Log Scores Comparison: SWFF vs SWπ



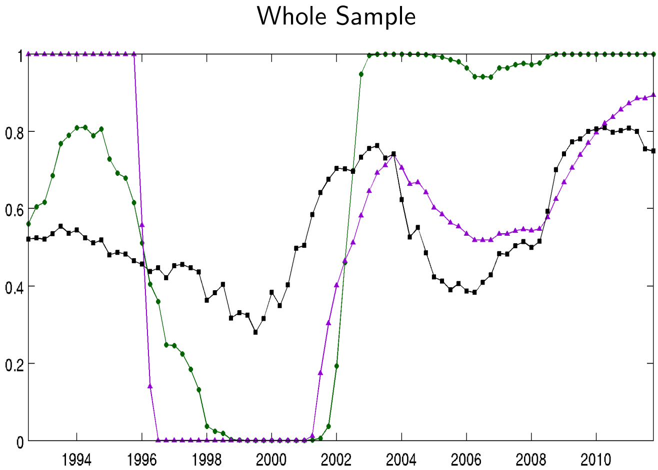
Static Pools – Recursive Estimation of λ

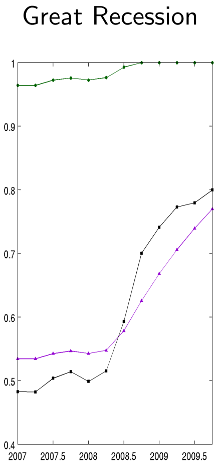


Dynamic Pools – Recursive Estimation of $\lambda_{T|T}$

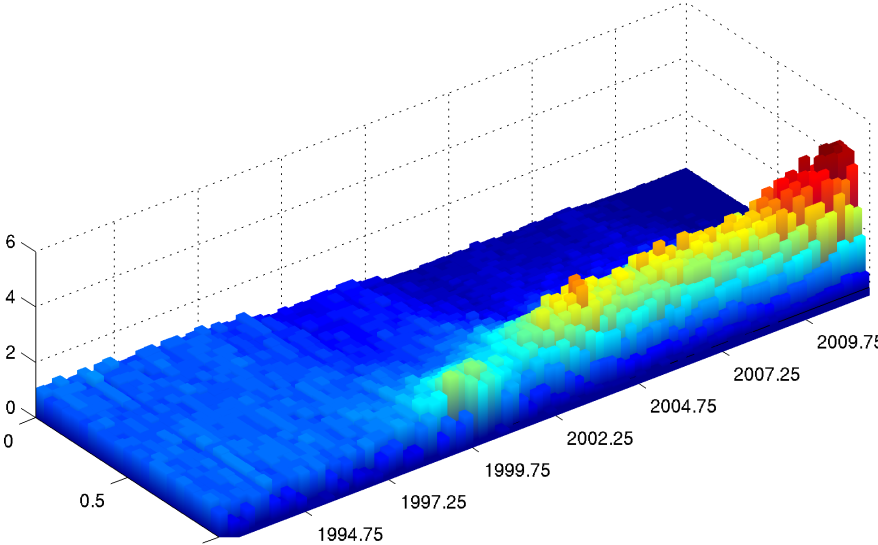


Dynamic, BMA, and Static Pool Weights in Real Time

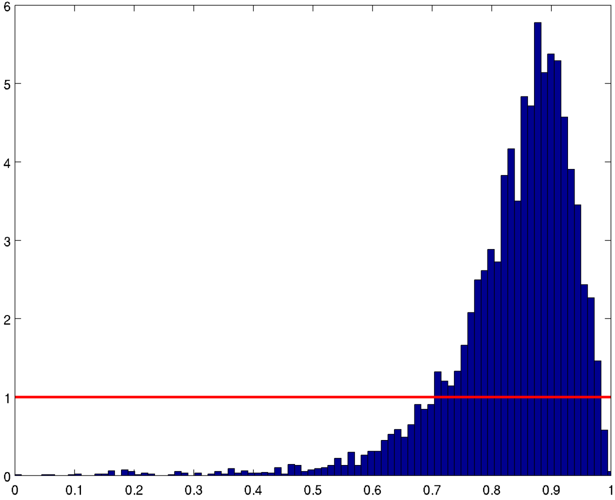




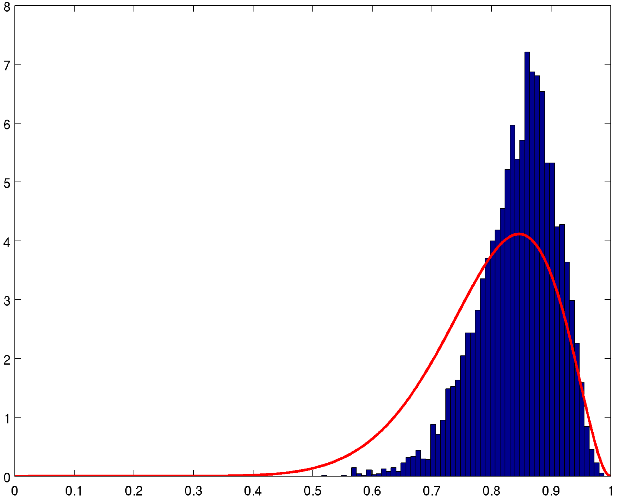
Recursive Posterior of ρ



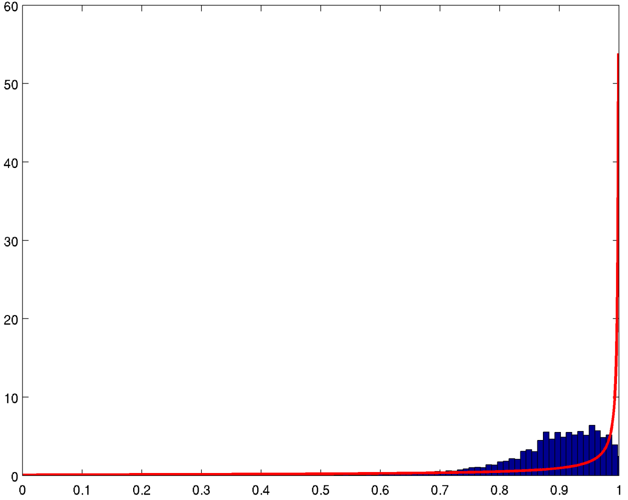
Prior/Posterior of ρ



Prior/Posterior of $\rho - \mathcal{B}(.8, .1)$ prior

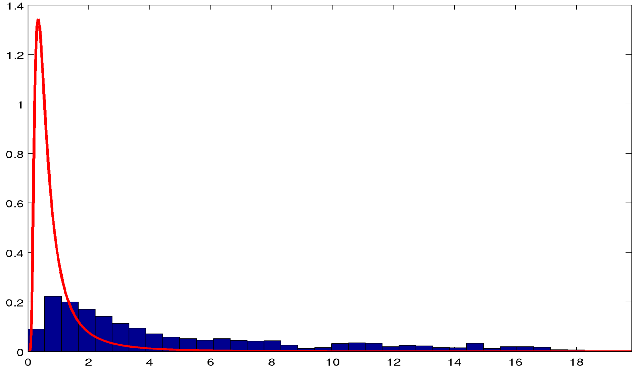


Prior/Posterior of ρ – $\mathcal{B}(.9, .2)$ prior



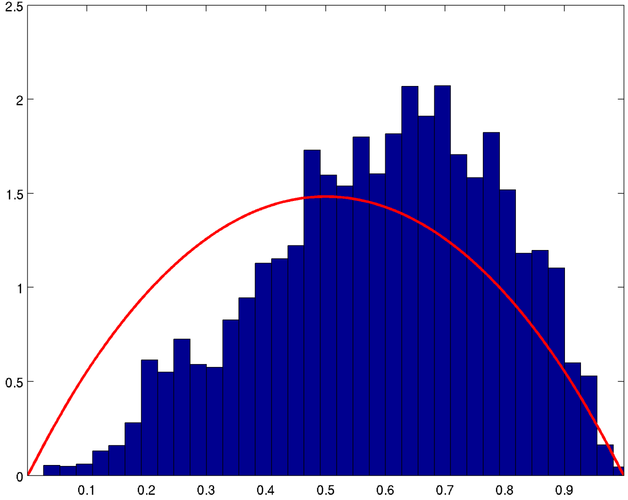
Prior/Posterior of σ^2 – $\mathcal{IG}(1,4)$ prior

$$\begin{aligned} x_t &= \mu + \rho x_{t-1} + \sigma\sqrt{1-\rho^2}\varepsilon_t,\ \varepsilon_t \sim iid\ N(0,1),\ x_0 \sim N(0,1), \\ \lambda_t &= \Phi(x_t) \end{aligned}$$

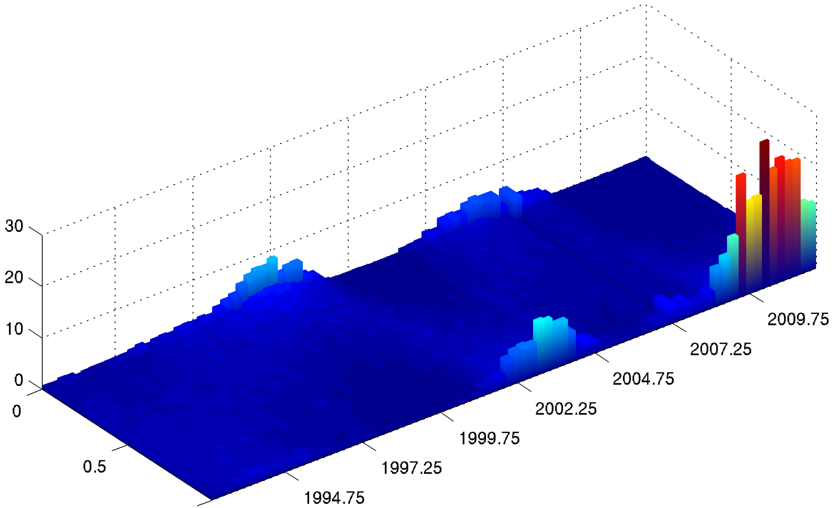


Marginal likelihood difference $\sigma^2 = 4\ -\ \sigma^2 = .5\ =\ \sim 5$ log points

Prior/Posterior of μ – in $\Phi(\mu)$ space

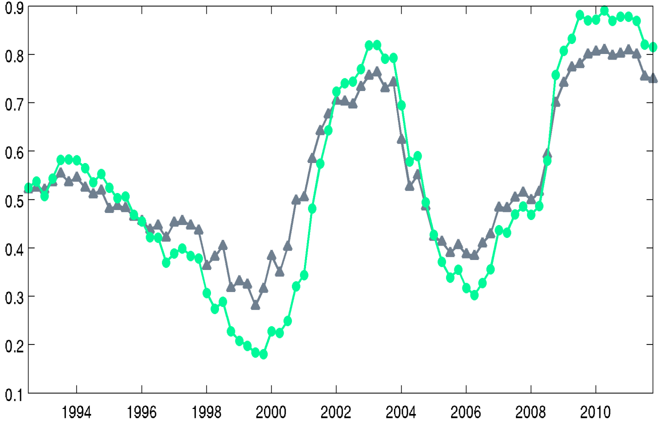


Dynamic Pools – Recursive Estimation of $\lambda_{T|T}$

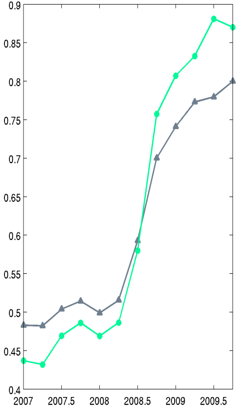


$\lambda_{T|T}$: Plain vs Estimated μ, σ DP

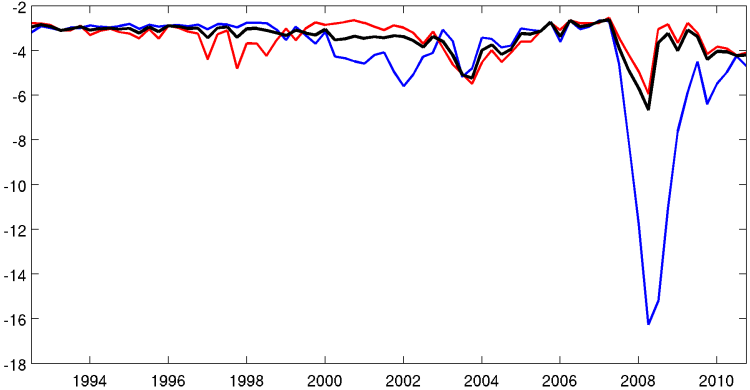
Whole Sample



Great Recession

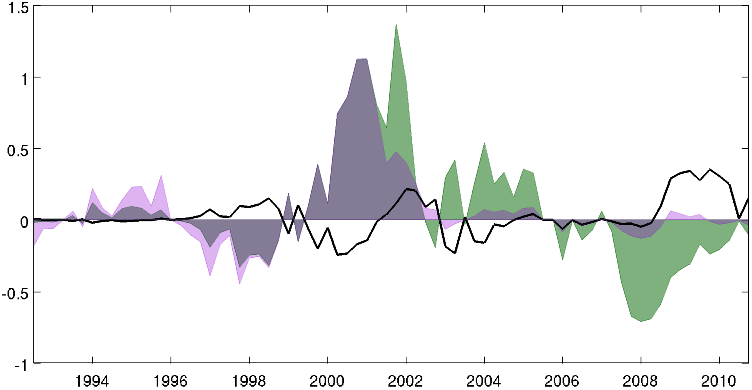


Log Scores Comparison: SWFF (*red*) vs SWπ (*blue*) vs DP (*black*)



$$p^{DP}(y_t|\lambda_{t-h|t-h}) = \lambda_{t|t-h} p(y_t|\mathcal{I}_{t-h}^{\text{SWFF}}, \text{SWFF}) + (1 - \lambda_{t|t-h}) p(y_t|\mathcal{I}_{t-h}^{\text{SW}\pi}, \text{SW}\pi)$$

Log score comparison: Dynamic vs BMA, Static Pool, and equal weights *(black line)*



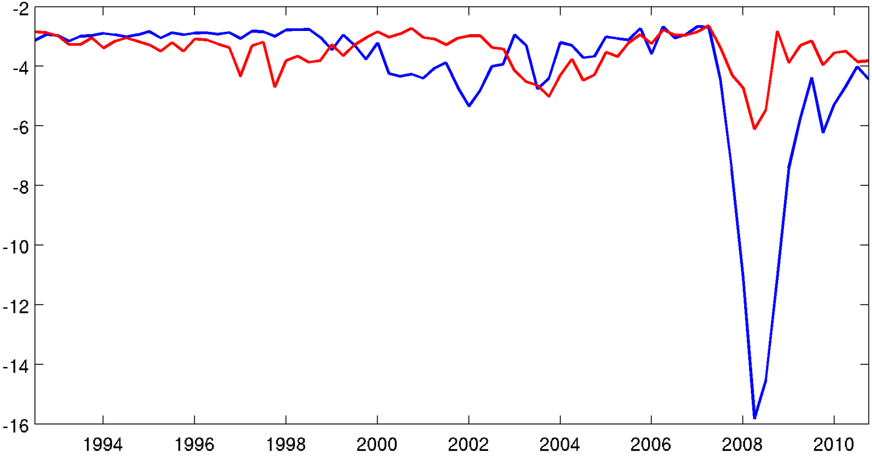
Cumulative Log Scores

	Log Score(DP)	Log Score(DP) - Log Score(Alt.)		
Specification	DP	EW	BMA	SP
$\rho \sim U(0,1),\ \mu = 0,\ \sigma^2 = 1$	-256.91	1.34	4.07	4.95
$\rho \sim U(0,1),$ $\mu \sim \mathcal{N}(0, \Phi^{-1}(.75)),\ \sigma^2 \sim \mathcal{IG}(1,4)$	-256.42	1.83	4.56	5.44
$\rho \sim \mathcal{B}(.8,.1),$ $\mu \sim \mathcal{N}(0, \Phi^{-1}(.75)),\ \sigma^2 \sim \mathcal{IG}(1,4)$	-256.43	1.82	4.55	5.43
$\rho \sim \mathcal{B}(.8,.1),\ \mu = 0,\ \sigma^2 \sim \mathcal{IG}(1,4)$	-255.97	2.28	5.01	5.89

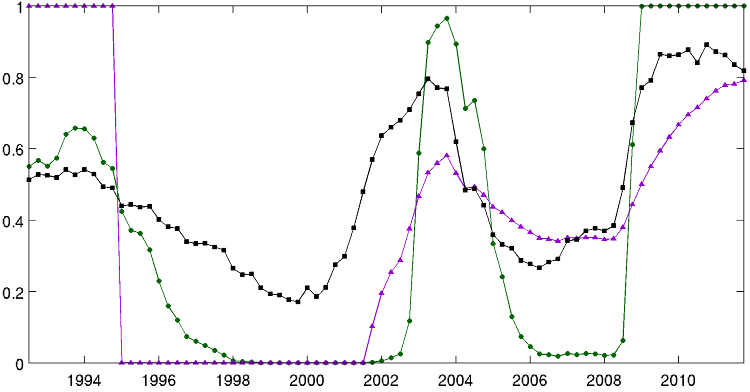
Conclusions

- Methodology:
  - There is evidence of time-variation in relative forecasting performance of different models over time (see also Amisano and Geweke 2013) $\rightarrow$ Dynamic Pools seem worth exploring.
- Substantive:
  - Evidence of time-variation in relative forecasting performance of DSGE models with and without financial frictions ...
  - ... yet no excuse for having ignored FF prior to the Great Recession.
  - Should use financial friction model now
  - Evidence in favor of non-linearities

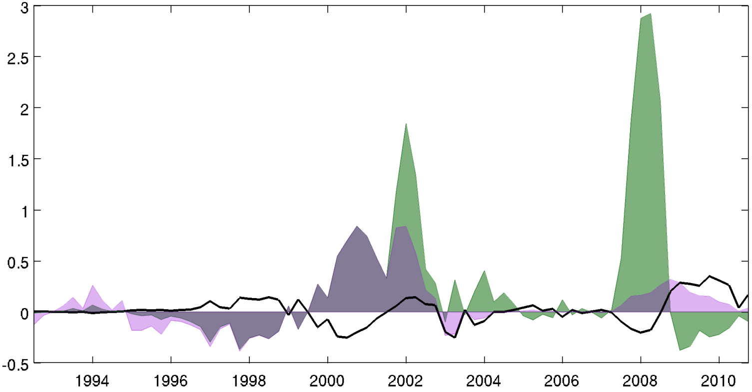
Log Scores Comparison: SWFF vs SWπ – No $t+1$



Dynamic, BMA, and Static Pool Weights in Real Time – No $t + 1$ Information



Log score comparison: Dynamic vs BMA, Static Pool, and equal weights *(black line)* – No $t + 1$ Information

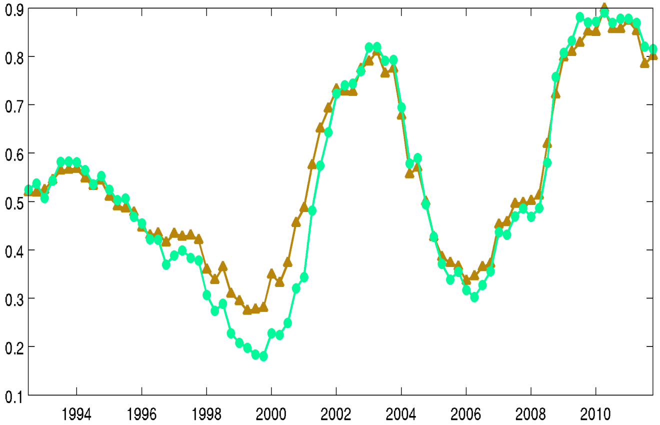


Cumulative Log Scores – No $t+1$ Information

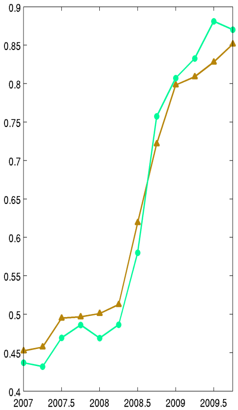
	Log Score(DP)	Log Score(DP) - Log Score(Alt.)		
Specification	DP	EW	BMA	SP
$\rho \sim U(0,1),\ \mu = 0,\ \sigma^2 = 1$	-259.43	0.99	16.14	5.24
$\rho \sim \mathcal{B}(.8,.1),$ $\mu \sim \mathcal{N}(0, \Phi^{-1}(.75)),\ \sigma^2 \sim \mathcal{IG}(1,4)$	-259.09	1.33	16.48	5.58
$\rho \sim \mathcal{B}(.8,.1),\ \mu = 0,\ \sigma^2 \sim \mathcal{IG}(1,4)$	-258.62	1.80	16.95	6.05

$\lambda_{T|T}$: $\mu = 0$ vs Estimated μ DP

Whole Sample



Great Recession



Prior/Posterior of μ