Combining Macroeconomic Models for Prediction

John Geweke
University of Technology Sydney

15th Australasian Macro Workshop

April 8, 2010

Outline

1. Optimal prediction pools
2. Models and data
3. Optimal pools for joint prediction
4. Optimal pools for individual time series
5. Conclusions and further research

Background:
Some of this work is joint with Gianni Amisaon, European Central Bank
Geweke and Amisano (2009),
Optimal Prediction Pools, ECB working paper 1017,
http://www.ecb.int/pub/pdf/scpwps/ecbwp1017.pdf
Geweke (2010),
Complete and Incomplete Econometric Models,
Princeton University Press

Optimal prediction pools: Econometric motivation

- There are often several models relevant for a decision
  - VAR's (Vector autoregression models)
  - DSGE's (Dynamic stochastic general equilibrium models)
  - DFM's (Dynamic factor models)
- Decision makers know that all of these models are simplifications
  - i.e., they are wrong.
- Bayesian and non-Bayesian methods assume one of the models is true.
- What happens if we remove this assumption?
- Geweke and Amisano (2009), Geweke (2010):
  - Detail on methodology
  - Application to asset returns
- This work: Optimal prediction pools of leading macroeconomic forecasting models

The setting

- Time series $\{\mathbf{y}_t\}$
- History $\mathbf{Y}_{t-1} = \{\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}\}$
- Prediction model A: a probability density $p\left(\mathbf{y}_t; \mathbf{Y}_{t-1}, A\right)$
  - Formal Bayesian approach:

$$\begin{aligned} p\left(\mathbf{y}_t; \mathbf{Y}^o_{t-1}, A\right) &= p\left(\mathbf{y}_t \mid \mathbf{Y}^o_{t-1}, A\right) \\ &= \int p\left(\mathbf{y}_t \mid \mathbf{Y}^o_{t-1}, \boldsymbol{\theta}_A, A\right) p\left(\boldsymbol{\theta}_A \mid \mathbf{Y}^o_{t-1}, A\right) d\boldsymbol{\theta}_A \end{aligned}$$

  - Common non-Bayesian approach:

$$\begin{aligned} \widehat{\boldsymbol{\theta}}^{t-1}_A &= f_{t-1}\left(\mathbf{Y}^o_{t-1}\right), \\ p\left(\mathbf{y}_t; \mathbf{Y}^o_{t-1}, A\right) &= p\left(\mathbf{y}_t \mid \mathbf{Y}^o_{t-1}, \widehat{\boldsymbol{\theta}}^{t-1}_A, A\right) \end{aligned}$$

- What matters: A produces a legitimate p.d.f. for $\mathbf{y}_t$, relying only on $\mathbf{Y}_{t-1}$ and A.

Log scoring

- Log predictive score:

$$LS\left(\mathbf{Y}_T^o, A\right) = \sum_{t=1}^{T} \log p\left(\mathbf{y}_t^o; \mathbf{Y}_{t-1}^o, A\right)$$

- Formal Bayesian approach

$$\begin{aligned} p\left(\mathbf{y}_t^o; \mathbf{Y}_{t-1}^o, A\right) &= p\left(\mathbf{y}_t^o \mid \mathbf{Y}_{t-1}^o, A\right), \\ LS\left(\mathbf{Y}_T^o, A\right) &= p\left(\mathbf{Y}_T^o \mid A\right) = \int p\left(\mathbf{Y}_T^o \mid \boldsymbol{\theta}_A, A\right) p\left(\boldsymbol{\theta}_A \mid A\right) d\boldsymbol{\theta}_A \end{aligned}$$

- Common non-Bayesian approach:

$$LS\left(\mathbf{Y}_T^o, A\right) = \sum_{t=1}^{T} \log p\left(\mathbf{y}_t^o \mid \mathbf{Y}_{t-1}^o, \widehat{\boldsymbol{\theta}}_A^{t-1}, A\right)$$

Prediction pools of multiple models

- In a prediction pool with n models the log predictive score function is

$$f_T(\mathbf{w}) = \sum_{t=1}^{T} \log \left[\sum_{i=1}^{n} w_i p\left(\mathbf{y}_t^o \mid \mathbf{Y}_{t-1}^o, A_i\right) \right]$$

  - where $\mathbf{w} = (w_1, \ldots, w_n)'$, $w_i \geq 0$ $(i = 1, \ldots, n)$ and $\sum_{i=1}^{n} w_i = 1$.

- For an ergodic data generating process D,

$$T^{-1} f_T(\mathbf{w}) \xrightarrow{a.s.} \lim_{T\to\infty} T^{-1} \int \log \left[\sum_{i=1}^{n} w_i p\left(\mathbf{y}_t \mid \mathbf{Y}_{t-1}, A_i\right) \right] \cdot p\left(\mathbf{Y}_T \middle| D\right) d\nu\left(\mathbf{Y}_T\right) = f(\mathbf{w}).$$

- Some short-hand:

$$p_{ti} = p\left(\mathbf{y}_t^o; \mathbf{Y}_{t-1}^o, A_i\right) \left(t = 1, \ldots, T; i = 1, \ldots, n\right)$$

Optimization

$$f_T(\mathbf{w}) = \sum_{t=1}^{T} \log\left[\sum_{i=1}^{n} w_i p\left(\mathbf{y}_t^o \mid \mathbf{Y}_{t-1}^o, A_i\right)\right] = \sum_{t=1}^{T} \log\left(\sum_{i=1}^{n} w_i p_{ti}\right)$$

- First derivative (after substituting $w_1 = 1 - \sum_{i=2}^{n} w_i$):

$$\partial f_T(\mathbf{w}) / \partial w_i = \sum_{t=1}^{T} \frac{p_{ti} - p_{t1}}{\sum_{j=1}^{n} w_j p_{tj}} \quad (i = 2, \ldots, n)$$

- Second derivative: $\partial^2 f_T(\mathbf{w}) / \partial w_i \partial w_j$

$$= -T^{-1} \sum_{t=1}^{T} \frac{\left(p_{ti} - p_{t1}\right)\left(p_{tj} - p_{t1}\right)}{\left[\sum_{k=1}^{n} w_k p_{tk}\right]^2} \quad (i, j = 2, \ldots, n)$$

- $f_T(\mathbf{w})$ is a concave function.
- Given the evaluations p_{ti} from the alternative prediction models and a sample, finding $\mathbf{w}_T^* = \arg\max_{\mathbf{w}} f_T(\mathbf{w})$ is a straightforward convex programming problem.

Population behavior

Review of model averaging and selection

- Recall that for each model A_j, $T^{-1} LS\left(\mathbf{Y}_T, A_j\right) \stackrel{a.s.}{\to}$

$$\lim_{T\to\infty} T^{-1} \int \left[\sum_{t=1}^{T} \log p\left(\mathbf{y}_t; \mathbf{Y}_{t-1}, A\right) \right] p\left(\mathbf{Y}_T | D\right) d\nu\left(\mathbf{Y}_T\right)$$

- Hence for all interesting pairs A_i and A_j,

$$LS\left(\mathbf{Y}_T, A_i\right) - LS\left(\mathbf{Y}_T, A_j\right) \stackrel{a.s.}{\to} \pm\infty.$$

- As a consequence
  - Bayesian procedures assign probability 1 to one model asymptotically
  - Non-Bayesian testing procedures select the same model asymptotically.
  - Asymptotically, these procedures all use a pseudo-true model with pseudo-true parameter values for prediction.
  - *This is the wrong answer under a log scoring rule.*

Population behavior

Limiting behavior of optimal prediction pools

- The population function

$$f(\mathbf{w}) = \lim_{T\to\infty} T^{-1} f_T(\mathbf{w}) = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \log\left(\sum_{i=1}^{n} w_i p_{ti}\right)$$

  is also concave.
- Define $\mathbf{w}^* = \arg\max f(\mathbf{w})$
  - $\mathbf{w}_T^* \stackrel{a.s.}{\rightarrow} \mathbf{w}^*$
  - Typically several elements of $\mathbf{w}^*$ are nonnegative...
  - Despite the fact that both Bayesian and non-Bayesian methods will use just one model in prediction asymptotically.
- What is the explanation?
  - Conventional Bayesian and non-Bayesian procedures assume $A_j = D$ for some $j = 1, \ldots, n$.
  - Optimal log scoring does not make this assumption.

Population behavior

What if one of the models were true?

- The population function is

$$f\left(\mathbf{w}\right)=\lim_{T\to\infty}T^{-1}f_T\left(\mathbf{w}\right)=\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\log\left(\sum_{i=1}^{n}w_ip_{ti}\right)$$

- Define $\mathbf{w}^*=\arg\max f\left(\mathbf{w}\right)$
- Proposition: If $A_1=D$, then $\mathbf{w}^*=(1,0,\ldots,0)$;
  - furthermore,

$$\frac{\partial f\left(\mathbf{w}\right)}{\partial w_j}\mid_{\mathbf{w}=\mathbf{w}^*}=0\quad\left(j=1,\ldots,m\right).$$

Overview of the models

- Vector autoregression (VAR)
- Dynamic stochastic general equilibrium model (DSGE)
- Dynamic factor model (DFM)
- In each case we used a variant of the model and a method of Bayesian inference representative of current practice at central banks.
- *Caveat*:
  - Work with several alternative variants is currently proceeding.
  - The initial results presented today may or may not be representative of results with these variants.

Data: An extension of Smets and Wouters (2007)

Quarterly U.S. data, 1951:I - 2009:I

1. Consumption: growth rate in per capita real consumption
2. Investment: growth rate in per capita real investment
3. Output: growth rate in per capita real GDP
4. Hours: log per capita weekly hours
5. Inflation: growth rate in GDP deflator
6. Real wage: growth rate in real wage
7. Interest rate: Federal Funds Rate

Additional series for DFM

1. Stock returns: Growth rate in S&P 500 index
2. Unemployment rate
3. Term premium: 10 year and 3 month bond rates spread
4. Risk premium: BAA and AAA corporate bond spread
5. Money growth: Growth rate in M2

Vector autoregression (VAR) model

- Conventional VAR with Minnesota priors
- VAR is in levels, predictive densities are for differences (except hours and interest rate)
- Full Bayesian inference using MCMC
- Four lags of each variable

Dynamic stochastic general equilibrium (DSGE) model

- Model described in Smets and Wouters, AER 2007
- DSGE model with nominal frictions: price and wage stickiness, monopolistic competition.
- "The marginal likelihood criterion, which captures the out-of-sample prediction performance, is used to test the [DSGE] model against standard and Bayesian VAR models. We find that the [DSGE] model has a fit comparable to that of Bayesian VAR models." (p. 587)
- Unit root structure: some exogenous driving variables are I(1), variables transformed to stationarity
- Seven structural shocks: total factor productivity, risk premium, investment specific tech shock, wage mark up, price mark up, exogenous government spending, monetary shock
- Bayesian inference with results based on posterior modal value of parameters (as in DYNARE)

Dynamic factor model (DFM)

- Model specification following Stock and Watson (2005, NBER working paper).
  - $k = 3$ common factors with VAR dynamics
  - $n = 12$ idiosyncratic terms with AR dynamics
- Structure:
  - $\underset{(12\times1)}{\mathbf{y}_t} = \Gamma \underset{(3\times1)}{\mathbf{f}_t} + \mathbf{v}_t$
  - $b_i(L)v_{it} = \varepsilon_{it}, i = 1, 2, ...12;$ lag length 2; $\varepsilon_t \overset{iid}{\sim} N(\mathbf{0}, diag(\sigma))$
  - $\mathbf{A}(L)\mathbf{f}_t = \boldsymbol{\eta}_t, \boldsymbol{\eta}_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_3);$ lag length 2
- Bayesian inference with proper priors
- Marginal predictive distribution for first 7 variables used for model pool

Log scores of individual models, 1966:I - 2009:I

VAR	-1042.5
DSGE	-1087.7
DFM	-1019.6

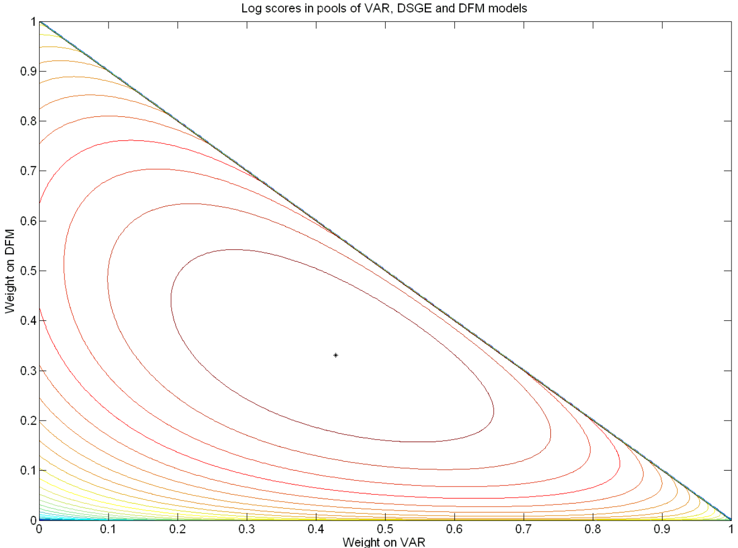
- Formal interpretation in VAR and DFM: Log marginal likelihood with
  - Prior and data 1951:I - 1965:IV constituting the prior distribution
  - Likelihood from the data 1966:I - 2009:I
- DSGE uses fixed parameter value (posterior mode) each quarter.

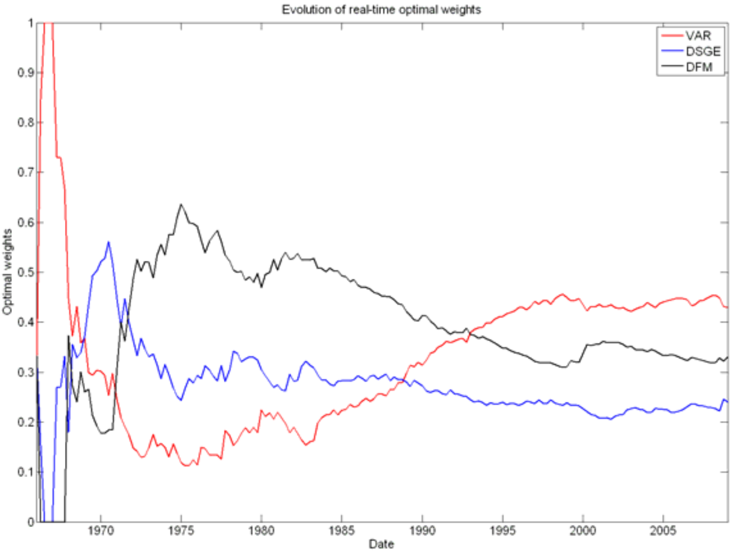
Optimal pool of models

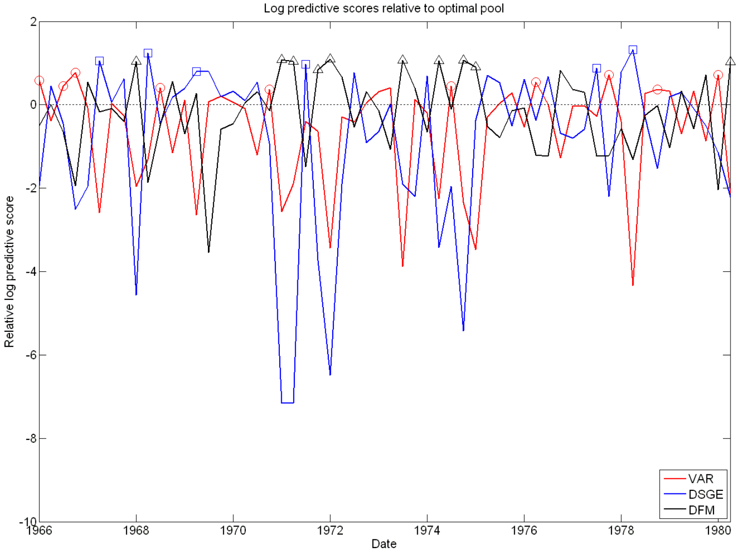
Model	VAR	DSGE	DFM
Log score	-1042.5	-1087.7	-1019.6
Weight	0.429	0.240	0.330
Value	14.6	10.0	40.8

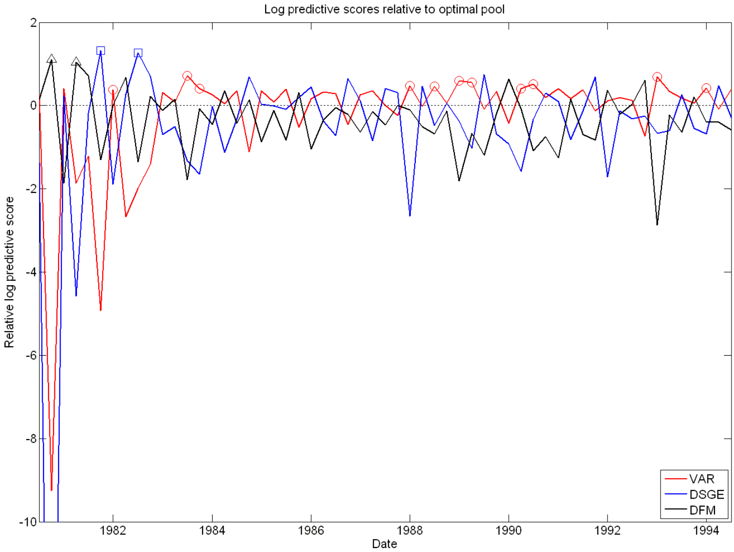
Log score of optimal pool: -974.9

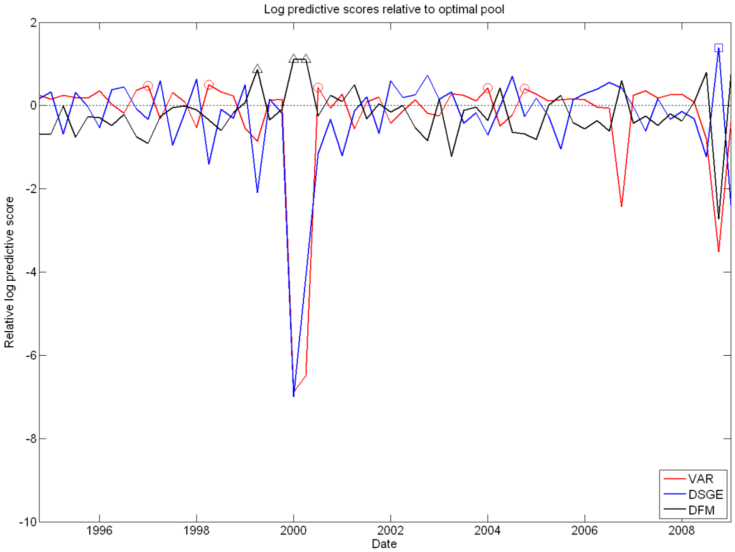
Log score of equally-weighted pool: -975.8


Log scores in pools of VAR, DSGE and DFM models
Weight on DFM
Weight on VAR
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Evolution of real-time optimal weights
VAR
DSGE
DFM
Optimal weights
Date
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1970
1975
1980
1985
1990
1995
2000
2005


Log predictive scores relative to optimal pool
Relative log predictive score
Date
2
0
-2
-4
-6
-8
-10
1966
1968
1970
1972
1974
1976
1978
1980
VAR
DSGE
DFM


Log predictive scores relative to optimal pool
Relative log predictive score
Date
2
0
-2
-4
-6
-8
-10
1982
1984
1986
1988
1990
1992
1994
VAR
DSGE
DFM


Log predictive scores relative to optimal pool
Relative log predictive score
Date
2
0
-2
-4
-6
-8
-10
1996
1998
2000
2002
2004
2006
2008
VAR
DSGE
DFM

Individual series: model weights in optimal pools

	VAR	DSGE	DFM
Hours	0.146	0.014	0.841
Interest rate	0.161	0.478	0.361
Inflation	0.341	0.659	0.000
Real GDP	0.135	0.000	0.865
Real consumption	0.356	0.211	0.434
Real investment	0.034	0.340	0.625
Real wage	0.451	0.549	0.000

Individual series: model values in optimal pools

	VAR	DSGE	DFM
Hours	0.069	0.004	2.779
Interest rate	0.305	6.109	17.21
Inflation	1.500	11.904	0.000
Real GDP	0.193	0.000	5.453
Real consumption	0.622	0.913	1.118
Real investment	0.006	1.286	1.819
Real wage	1.791	3.378	0.000

Summary

- Optimal pooling:
  - Does not assume one of the models is true
  - Weights are very different from Bayesian posterior probabilities
  - Many more properties in Geweke and Amisano (2009), Geweke(2010)
- In the optimal pool of VAR, DSGE and DFM models
  - All three models have positive weight and value
  - VAR has the highest weight, DFM the greatest value, and DSGE the lowest weight and value
  - For marginal predictive densities (individual series) results are varied
  - Strong indication that no model is (close to) DGP
  - Consistent with the observation that all three models are used by central banks despite the fact that posterior odds overwhelmingly favors DFM

Further research

- The application:
  - Interpretation of results
  - Variants on each of the three models
  - Variants on methods of inference
  - Data from other countries
- Optimal pooling:
  - Nonlinear pools
  - Alternative utility functions