Forecast Accuracy and Economic Gains from Bayesian Model Averaging using Time Varying Weights *

Herman K. van Dijk

(Econometric Institute & Tinbergen Institute,
Erasmus University Rotterdam)

Joint work with: Lennart Hoogerheide (EUR)
Richard Kleijn (PGGM)
Francesco Ravazzolo (Norges Bank)
Marno Verbeek (EUR)

* TI paper 2009-061 (www.tinbergen.nl), *Journal of Forecasting 2009*

Outline:

- Some literature

- Forecast combination schemes:

  - scheme 1 BMA: Bayesian Model Averaging

  - Combination schemes using estimated regression coefficients as model weights:

    scheme 2 LIN: Model weights from Ordinary Least Squares in a linear model

    scheme 3 TVW: Time-varying weights

    scheme 4 RTVW: Robust time-varying weights

- Applications:
  - financial: S&P500 monthly returns
  - macro: US quarterly real GDP growth

Literature on forecast combinations:

Since "Combination of Forecasts" by Bates & Granger (1969, *Operational Research Quarterly*) a huge number of publications has appeared.

For a wide range of time series processes, forecast combinations have appeared to perform better than forecasts based on single models.

Diebold and Pauly (1987, *Journal of Forecasting*) **regression based approach with time varying parameters.**

Some of the recent publications:

Terui & Van Dijk (2002, *International J of Forecasting*), "Combined forecasts from linear and nonlinear time series models", generalize the least squares model weights by reformulating the linear regression model as a state space specification,where the weights are assumed to follow a *random walk process*.

Literature on forecast combinations: <u>Some recent publications:</u>

Stock & Watson (2004, *J of Forecasting*), "Combination Forecasts of Output Growth in a Seven-country Data Set".

Hendry & Clements (2004, *Econometric Reviews*), "Pooling of Forecasts".

Timmermann (2006, *Handbook of Economic Forecasting*), "Forecast Combinations".

Stock & Watson (2004) and Timmermann (2006) compute model weights using the *inverse mean square prediction error* (MSPE) over a set of the most recent observations.

Hendry & Clements (2004) and Timmermann (2006) show that *simple* combinations (e.g. averages) often give better performance than more *sophisticated* combination schemes (with weights depending on the full covariance matrix of forecast errors).

Geweke & Whiteman (2006, *Handbook of Economic Forecasting), "*Bayesian Forecasting".

Guidolin & Timmermann (2009, *J of Econometrics*, forthcoming), "Forecasts of US Short-term Interest Rates: A Flexible Forecast Combination Approach"

Strachan & Van Dijk (2008), "Bayesian Averaging over Many Dynamic Model Structures with Evidence on the Great Ratios and Liquidity Trap Risk", Tinbergen Institute report 2008-096/4.

Geweke and Amisano, Optimal Prediction Pools, 2008.

Geweke & Whiteman (2006) apply BMA using *predictive* likelihoods instead of *marginal* likelihoods.

Strachan & Van Dijk (2008) compute *impulse response paths and effects of policy measures* using BMA in the context of a large set of VAR models.

Guidolin & Timmermann (2009) propose model weights having *regime switching dynamics*.
Geweke and Amisano (2008), propose *prediction pools* evaluated using log *predictive scoring rule*.

We propose: 3 forecast combination schemes that simultaneously allow for:

[1] parameter uncertainty

[2] model uncertainty

[3] time varying model weights

These approaches can be considered Bayesian extensions of the combination scheme of Terui & Van Dijk (2002).

We compare the performance of the proposed methods with Bayesian Model Averaging (BMA).

Scheme 1 BMA: **Bayesian Model Averaging:**

Compute predictive density of y_{T+1} (conditional upon D_T, data up to time T):

$$p(y_{T+1} \mid D_T) = \sum_{i=1}^{n} p(y_{T+1} \mid D_T,\ m_i)\Pr[m_i \mid D_T]$$

with: n = number of individual models

$p(y_{T+1} \mid D_T,\ m_i)$ = conditional predictive density given model m_i

$\Pr[m_i \mid D_T]$ = posterior probability of model m_i

The conditional predictive density given model m_i is:

$$p(y_{T+1} \mid D_T,\ m_i) = \int p(y_{T+1} \mid D_T,\ m_i, \theta_i)\, p(\theta_i \mid D_T,\ m_i)\, d\theta_i$$

with $p(\theta_i \mid D_T,\ m_i)$ the posterior density of parameters θ_i in model m_i.

Scheme 1 BMA: Bayesian Model Averaging (continued)

The posterior probability of model m_i is:

$$\Pr[m_i \mid D_T] = \frac{p(y_{1:T} \mid m_i)\Pr[m_i]}{\sum_{j=1}^{n} p(y_{1:T} \mid m_j)\Pr[m_j]}$$

with $\Pr[m_i]$ the prior probability for model m_i, and $p(y_{1:T} \mid m_i)$ the marginal likelihood:

$$p(y_{1:T} \mid m_i) = \int p(y_{1:T} \mid m_i, \theta_i)\, p(\theta_i \mid m_i)\, d\theta_i$$

with $p(\theta_i \mid m_i)$ the prior density for parameters θ_i in model m_i.

Chib (1995, *JASA*), “Marginal Likelihood from the Gibbs Output”

Ardia, Hoogerheide & Van Dijk (2009), “To Bridge, to Warp or to Wrap? A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihoods.” Tinbergen institute report 09-017.

Scheme 1 BMA: Bayesian Model Averaging (continued)

We follow Geweke & Whiteman (2006), and use *predictive* likelihood rather than *marginal* likelihood:

$$\Pr[m_i \mid D_T] = \frac{p(y_{(k+1):T} \mid m_i, D_k)\Pr[m_i]}{\sum_{j=1}^{n} p(y_{(k+1):T} \mid m_j, D_k)\Pr[m_j]}$$

with 'initial period' of k=12 (months), and

$$p(y_{(k+1):T} \mid m_i, D_k) = \prod_{t=k+1}^{T} p(y_t \mid m_i, D_{t-1})$$

The densities $p(y_t \mid m_i, D_{t-1})$ are evaluated as follows:

(1) parameters θ_i are simulated from the conditional distribution on D_{t-1}.

(2) draws y_t are simulated conditionally on the θ_i draws and D_{t-1}.

(3) a kernel smoothing technique is used to estimate the density of y_t in model m_i at its realized value.

Scheme 1 BMA: **Bayesian Model Averaging (continued)**

In all models, we specify uninformative proper priors for the parameters θ_i.

The use of predictive likelihoods rather than marginal likelihoods helps us to avoid the inference problems due to the Bartlett paradox.

Forecast combination schemes using estimated regression coefficients as model weights:

The three proposed forecast combination schemes estimate the weights w_i of the models m_i $(i = 1,\ldots,n)$ in regression form.

We assume that the data y_t satisfy the linear equation:

$$y_t = w_0 + \sum_{i=1}^{n} w_i \, y_{t,i} + u_t \qquad u_t \sim IID(0,\sigma^2) \qquad t = 1,2,\ldots,T$$

where $y_{t,i}$ has the predictive density $p(y_t \mid D_{t-1}, m_i)$.

Differences with BMA: - a constant term w_0 is added

- no restriction that weights w_i ≥ 0 or $\sum_{i=1}^{n} w_i = 1$

$\Rightarrow$ weights w_i (i=1,…,n) can not be interpreted as model probabilities

---

Granger & Ramanathan (1984, *J of Forecasting*): constant term must be added to avoid biased forecasts, often leading to more accurate forecasts.

Forecast combination schemes using estimated regression coefficients as model weights (continued):

$$y_t = w_0 + \sum_{i=1}^{n} w_i \, y_{t,i} + u_t \qquad u_t \sim IID(0, \sigma^2) \qquad t = 1, 2, \ldots, T$$

with $y_{t,i} \sim p(y_t \mid D_{t-1}, m_i)$.

We propose three novel sampling algorithms for simulating model weight vectors $w = (w_0, w_1, \ldots, w_n)$ given data $y_{1:T}$ and predictive densities $p(y_t \mid D_{t-1}, m_i)$:

scheme 2 LIN: Model weights from OLS in a linear model

scheme 3 TVW: Time-varying weights

scheme 4 RTVW: Robust time-varying weights

scheme 2 LIN: Model weights from OLS in a linear model

$$y_t = w_0 + \sum_{i=1}^{n} w_i \, y_{t,i} + u_t \qquad u_t \sim IID(0,\sigma^2) \qquad \text{with } y_{t,i} \sim p(y_t \mid D_{t-1}, m_i).$$

[a] Generate a set of S model weights w^s (s = 1,…,S) by:

(i) simulating independently S sets of T x n draws $y_{t,i}^s$ from the predictive densities $p(y_t \mid D_{t-1}, m_i)$ $(t = 1,...,T; i = 1,...,n)$

(ii) estimating w^s as OLS estimate in: $y_t = w_0 + \sum_{i=1}^{n} w_i \, y_{t,i}^s + u_t^s$

[b] Use the model weights w^s to combine draws $y_{T+1,i}^s$ from predictive densities $p(y_{T+1} \mid D_T, m_i)$ into *'combined draws'* $\tilde{y}_{T+1}^s$:

$$\tilde{y}_{T+1}^s = w_0^s + \sum_{i=1}^{n} w_i^s \, y_{T+1,i}^s$$

The median of $\tilde{y}_{T+1}^s$ $(s = 1,...,S)$ is our point forecast $\hat{y}_{T+1}$ of y_{T+1}.

scheme 2 LIN: Model weights from OLS (continued)

OLS estimate in: $$y_t = w_0 + \sum_{i=1}^{n} w_i \, y_{t,i}^s + u_t^s$$

Note: - OLS is interpreted as posterior mean under flat prior.

- OLS estimator's frequentist property of *consistency* (for consistency no requirement of normality, homoskedastiocity, absence of serial correlation). In combination with taking median of $\tilde{y}_{T+1}^s$, this implies that the scheme is robust against the distribution of u_t^s.

- Scheme 2 can be considered as an extension of Granger & Ramanathan (1984) who combine point forecasts using weights that minimize a square loss function, to making use of Bayesian density forecasts.

(Simple geometric interpretation: Model weights minimize distance between vector of observed values $y_{1:T}$ and the space spanned by the constant vector and vectors of ‘predicted’ values $y_{1:T,i}^s$.)

scheme 2 LIN: Model weights from OLS (continued), Interpretation

The *'combined draws'* $\tilde{y}_{T+1}^{s}$

$$\tilde{y}_{T+1}^{s} = w_0^s + \sum_{i=1}^{n} w_i^s \, y_{T+1,i}^{s}$$

are interpreted as draws from a *'shrunk' predictive density* that aims at describing the *central part* of the predictive density, taking into account the parameter uncertainty and model uncertainty.

We compute the point forecast as the median of the *'combined draws'* $\tilde{y}_{T+1}^{s}$, where the median is preferred over the mean, because it is more robust to extreme draws.

scheme 3 TVW: Time-varying weights

Idea behind forecast combination: complementary roles of different models in approximating the data generating process.

These complementary roles in approximating the data generating process may differ over time $\Rightarrow$ allow the model weights to change over time:

$$y_t = w_{t,0} + \sum_{i=1}^{n} w_{t,i}\, y_{t,i} + u_t \qquad u_t \sim IID(0,\sigma^2) \qquad \text{with } \; y_{t,i} \sim p(y_t \mid D_{t-1}, m_i).$$

As Terui & Van Dijk (2002), we assume that the $w_t = (w_{t,0}, w_{t,1}, \ldots, w_{t,n})'$ $(t = 1,\ldots,n)$ evolve over time as:

$$w_t = w_{t-1} + \xi_t \qquad \xi_t \sim N(0,\Sigma)$$

We assume Σ to be diagonal, making the scheme computationally easier.
(This does not rule out that *a posteriori* there will be coinciding (large) changes of model weights; merely that this is not imposed *a priori*.
Still, we intend to analyze the extension to non-diagonal Σ in future research.)

scheme 3 TVW: Time-varying weights (continued)

A Kalman filter algorithm is used to iteratively update the subsequent model weights w_{t+1}^{s} (t=1,…,T+1) in the model

$$y_t = w_{t,0}^{s} + \sum_{i=1}^{n} w_{t,i}^{s}\, y_{t,i}^{s} + u_t^{s} \qquad u_t^{s} \sim N(0, \sigma^2)$$

We *fix* the values of σ^2 and the diagonal elements of Σ. A Bayesian can interpret these assumptions as having priors on σ^2 and Σ with 0 variances.*

For each *s* the parameters σ^2 and Σ could also be estimated by maximum likelihood or MCMC methods, but we discard this to reduce computational time.

---

* In the financial application (with *n* = 4 models) we set σ^2 = OLS estimate, diag(Σ) = (0.1, 0.01, …, 0.01) to have (small) *signal-to-noise ratios* in [0.005,0.01]. For robustness we have tried different σ^2, Σ with signal-to-noise ratios ranging from 0.0001 to 0.1, all resulting in qualitatively equal results.

scheme 3 TVW: Time-varying weights (continued)

The model weights w_t^s incorporate a trade-off between minimizing the differences between observed values $y_{1:T}$ and linear combinations of 'predicted' values $y_{1:T,i}^s$ $(i = 1,\ldots,n)$, and constructing a 'smooth' path of weights w_t^s over time.

As in scheme 2, we use the model weights w_{T+1}^s to combine draws $y_{T+1,i}^s$ from predictive densities $p(y_{T+1} \mid D_T, m_i)$ into *'combined draws'* $\tilde{y}_{T+1}^s$:

$$\tilde{y}_{T+1}^s = w_{T+1,0}^s + \sum_{i=1}^{n} w_{T+1,i}^s \, y_{T+1,i}^s$$

The median of $\tilde{y}_{T+1}^s$ $(s = 1,\ldots,S)$ is our point forecast $\hat{y}_{T+1}$ of y_{T+1}.

scheme 4 RTVW: Robust time-varying weights

Recently, a new specification has been developed that makes parameter estimation in case of instability over time more robust to prior assumptions, see e.g. Giordani & Villani (2008) and Groen, Paap & Ravazzolo (2009).

We extend scheme 3 of time-varying model weights following this reasoning:

$$w_t = w_{t-1} + k_t \odot \xi_t \qquad \xi_t \sim N(0, \Sigma)$$

with k_t = $(k_{t,0}, k_{t,1}, \ldots, k_{t,n})'$ where each element $k_{t,i}$ of the vector k_t is an unobserved 0/1 variable with $\Pr[k_{t,i} = 1] = \pi_i$.

The Hadamard product $\odot$ refers to element-by-element multiplication. Σ is again restricted to be a diagonal matrix.

Giordani & Villani (2008), “Forecasting macroeconomic time series with locally adaptive signal extraction”. Working paper.

Groen, Paap & Ravazzolo (2009), “Real-time inflation forecasting in a changing world.” Working paper.

scheme 4 RTVW: Robust time-varying weights (continued)

The model

$$y_t = w_{t,0}^s + \sum_{i=1}^{n} w_{t,i}^s \, y_{t,i}^s + u_t^s \qquad u_t^s \sim N(0, \sigma^2)$$

$$w_t^s = w_{t-1}^s + k_t^s \odot \xi_t^s \qquad \xi_t^s \sim N(0, \Sigma)$$

is estimated following Gerlach, Carter & Kohn (2000, *JASA*),
"Efficient Bayesian inference for dynamic mixture models" :

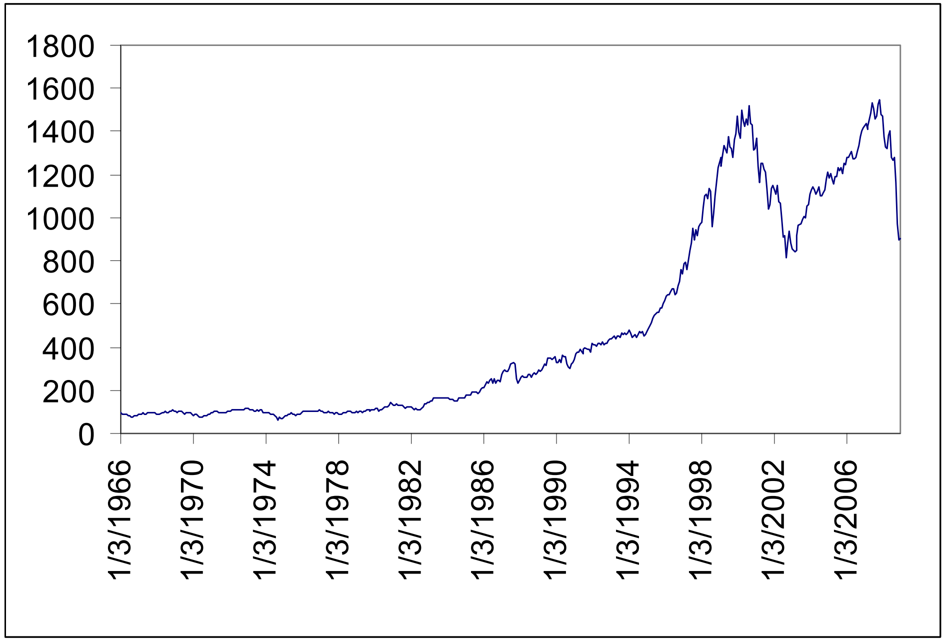
- deriving the posterior density of k_t^s conditional on σ^2, Σ (but not on w_t^s)
- then applying the Kalman Filter to estimate the latent factors w_t^s

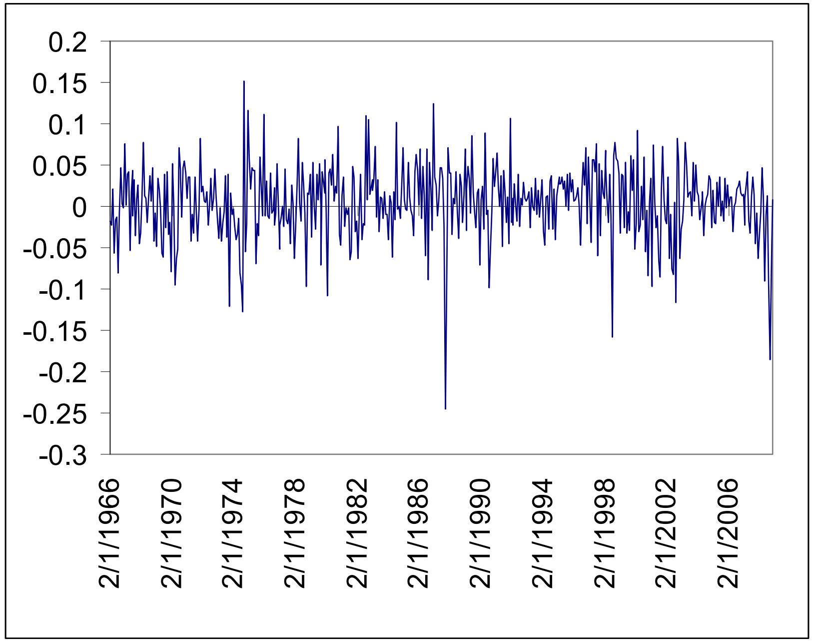
We set σ^2 and Σ to the same fixed values as for scheme 3.

Financial application: forecasting monthly S&P 500 returns

Data: continuously compounded monthly return on S&P 500 index in excess of 1-month T-Bill rate

Period: January 1966 - December 2008 (516 observations)





Bear market periods:
- burst of the internet bubble in 2001-2003
- recent financial crisis in 2nd part of 2007 & 2008

Financial application: forecasting S&P 500 returns (continued)

We compare our 4 forecast combination schemes:
- forecasting performance
- economic gains

We use *n* = 4 individual models:

Model 1 Leading Indicator (LI): linear model with lagged financial and macroeconomic variables (taking into account the typical publication lag of macroeconomic variables)

Model 2: Halloween Indicator (HI): linear regression model with a constant and a dummy for November-April. ("Sell in May and go away" of Bouman & Jacobsen (2002, *AER*))

Model 3: Stochastic Volatility (SV) with time-varying mean and volatility

Model 4: Robust Stochastic Volatility (RSV) with time-varying mean & vol.

Financial application: forecasting S&P 500 returns (continued)

Model 1 Leading Indicator (LI): explanatory variables (1-month lag):

- S&P 500 index dividend yield (ratio of dividends over previous 12 months and current stock price)
- 3-month T-Bill rate, monthly change in 3-month T-bill rate
- term spread (difference between 10-year T-bond rate & 3-month T-bill rate)
- credit spread (difference between Moody's Baa and Aaa yields)
- yield spread (difference between Federal funds rate and 3-month T-bill rate)
- annual inflation rate (producer price index (PPI) for finished goods) **
- annual growth rate of industrial production **
- annual growth rate of monetary base measure M1 **

** *2- month lag (publication lag)*

Financial application: forecasting S&P 500 returns (continued)

Model 3: Stochastic Volatility (SV) with time-varying mean and vol.:

$$r_t = \mu_t + \sigma_t u_t \qquad u_t \sim N(0,1)$$

$$\mu_t = \mu_{t-1} + \xi_{1,t} \qquad \xi_{1,t} \sim N(0,\tau_1^2)$$

$$\log(\sigma_t^2) = \log(\sigma_{t-1}^2) + \xi_{2,t}, \qquad \xi_{2,t} \sim N(0,\tau_2^2)$$

Model 4: Robust SV (RSV) with time-varying mean & vol.:

$$r_t = \mu_t + \sigma_t u_t \qquad u_t \sim N(0,1)$$

$$\mu_t = \mu_{t-1} + K_{1,t}\,\xi_{1,t} \qquad \xi_{1,t} \sim N(0,\tau_1^2)$$

$$\log(\sigma_t^2) = \log(\sigma_{t-1}^2) + K_{2,t}\,\xi_{2,t}, \qquad \xi_{2,t} \sim N(0,\tau_2^2)$$

$K_{1,t}$, $K_{2,t}$ $(t = 1,\ldots,T)$ are unobserved variables with

$$\Pr[K_{1,t} = 1] = \pi_{1,RSV}$$

$$\Pr[K_{2,t} = 1] = \pi_{2,RSV}$$

For Bayesian estimation of SV, RSV models: Giordani, Kohn & Van Dijk (2007, *J of Econometrics*), “A unified approach to nonlinearity, outliers & structural breaks.”

Financial application: forecasting S&P 500 returns (continued)

We compare 8 approaches:
- models 1, 2, 3, 4
- forecast schemes 1, 2, 3, 4

We evaluate:

- statistical accuracy:
  - root mean square prediction error (RMSPE)
  - correctly predicted percentage of sign (Sign Ratio)

- economic gains: returns for active short-term investment exercise (investment horizon of 1 month), with portfolio consisting of S&P500 and riskfree bonds only:
  - ex post annualized mean portfolio return
  - annualized standard deviation,
  - annualized Sharpe ratio
  - total utility.

Financial application: forecasting S&P 500 returns (continued)

Active short-term investment exercise (investment horizon of 1 month):

At start of each month T +1, investor decides upon fraction pw_{T+1} of her portfolio to be invested in stocks, based upon density forecast of excess stock return r_{T+1}. Wealth W_{T+1} at end of month T+1 will be:

$$W_{T+1} = W_T\Big((1 - pw_{T+1})\exp(r_{f,T+1}) + pw_{T+1}\exp(r_{f,T+1} + r_{T+1})\Big).$$

Investor chooses pw_{T+1} to maximize expected utility

$$\max_{pw_{T+1}} E[u(W_{T+1}) \mid D_T] \; = \; \max_{pw_{T+1}} \int u(W_{T+1}) p(r_{T+1} \mid D_T)\, dr_{T+1}.$$

We assume power utility function with coefficient of relative risk aversion γ:

$$u(W_{T+1}) = \frac{W_{T+1}^{1-\gamma}}{1-\gamma}, \qquad \gamma > 1.$$

Financial application: forecasting S&P 500 returns (continued)

Without loss of generality we set initial wealth equal to one, $W_T = 1$.

We approximate expected utility $E[u(W_{T+1}) \mid D_T]$ = $\int u(W_{T+1}) p(r_{T+1} \mid D_T)\, dr_{T+1}$:

(i) generating G draws r_{T+1}^g ($g = 1,\dots,G$) from predictive density $p(r_{T+1} \mid D_T)$

(ii) computing: $\hat{E}[u(W_{T+1}) \mid D_T]$ =

$$\frac{1}{G}\sum_{g=1}^{G}\frac{1}{1-\gamma}\Big((1-pw_{T+1})\exp(r_{f,T+1}) + pw_{T+1}\exp(r_{f,T+1} + r_{T+1}^{g})\Big)^{1-\gamma}$$

Then we find pw_{T+1} maximizing $\hat{E}[u(W_{T+1}) \mid D_T]$ using a numerical optimization method.

Note: We do not allow for short-sales or leveraging, i.e. constraining pw_{T+1} to be in the [0,1] interval (see Barberis (2000, *J of Finance*)).

Financial application: forecasting S&P 500 returns (continued)

Utility levels are used to compare the forecast approaches: realized utility levels are computed by substituting the realized return of the portfolios.

Total utility is then the sum of $u(W_{T+1})$ across all T^* investment periods $T = T_0, ..., T_0 + T^* - 1$, with first investment decision made at end of period T_0.

In order to compare alternative strategies we compute the multiplication factor of wealth that would equate their average utilities.
For example, suppose we compare two strategies A and B, providing wealth $W_{A,T+1}$, $W_{B,T+1}$ at time T +1. Then we determine Δ such that

$$\sum_{T=T_0}^{T_0+T^*-1} u(W_{A,T+1}) = \sum_{T=T_0}^{T_0+T^*-1} u(W_{B,T+1} / \exp(\Delta))$$

Following Fleming, Kirby & Ostdiek (2001, *J of Finance*), we interpret Δ as the maximum performance fee the investor would be willing to pay to switch from strategy A to strategy B.

Financial application: forecasting S&P 500 returns (continued)

For Δ it holds that *under a power utility specification:*

$$\Delta_{\text{A versus B}} = \Delta_{\text{A versus C}} - \Delta_{\text{B versus C}}$$

That is, the performance fee an investor is willing to pay to switch from strategy A to strategy B can also be computed as the difference between performance fees of these strategies with respect to a benchmark strategy C.

We consider 3 static benchmark strategies:

[i] holding stocks only (Δ_s)
[ii] holding a portfolio consisting of 50% stocks, 50% bonds (Δ_m),
[iii] holding bonds only (Δ_b).

Finally, the portfolio weights change every month, so the portfolio must be rebalanced accordingly. Hence, transaction costs play a non-trivial role. Therefore, we also consider the results under transaction costs of 0.1%.

Financial application: forecasting S&P 500 returns (continued)

Empirical results

Active investment strategies are implemented for Jan 1987 - Dec 2008, involving T^* = 264 one month ahead forecasts of excess stock return.

Individual models are estimated recursively using an expanding window. The initial 12 predictions for each individual model are used as training period for combination schemes and making the first combined prediction.

Statistical accuracy:

	model				combination scheme			
	1	2	3	4	1	2	3	4
	LI	HI	SV	RSV	BMA	LIN	TVW	RTVW
RMSPE	4.618	4.478	4.509	4.470	4.500	4.514	4.484	4.485
Sign ratio	0.527	0.549	0.614	0.598	0.587	0.610	0.602	0.598

⇒ Conclusion: performance of models and combination schemes similar.
(RSV, SV models best at RMSPE, sign ratio; but differences small)

Financial application: forecasting S&P 500 returns (continued)

The investment strategies are implemented for a level of relative risk aversion of γ = 6 (γ = 4 or γ = 8 results in qualitatively similar results).

Economic gains: (without transaction costs)

	model				combination scheme			
	1	2	3	4	1	2	3	4
	LI	HI	SV	RSV	BMA	LIN	TVW	RTVW
mean return	4.708	4.741	4.812	4.657	4.701	5.177	5.021	5.785
st dev return	0.794	0.769	1.139	0.614	0.739	4.356	1.332	3.062
Sharpe ratio	0.110	0.156	0.168	0.060	0.108	0.128	0.301	0.380
realized utility	-51.77	-51.76	-51.75	-51.79	-51.77	-51.73	-51.70	-51.56
Δ_s	285.5	288.7	295.2	277.9	283.8	304.3	317.1	381.3
Δ_m	-63.71	-60.49	-54.03	-71.29	-65.42	-44.95	-32.09	32.10
Δ_b	11.46	14.68	21.14	3.876	9.748	30.22	43.07	107.3

$\Rightarrow$ RTVW combination scheme best: highest mean return, Sharpe ratio, performance fees; highest (least negative) utility.

In fact, only RTVW has Δ_m > 0: only strategy beating 50% stock, 50% bond.

Financial application: forecasting S&P 500 returns (continued)

Empirical results

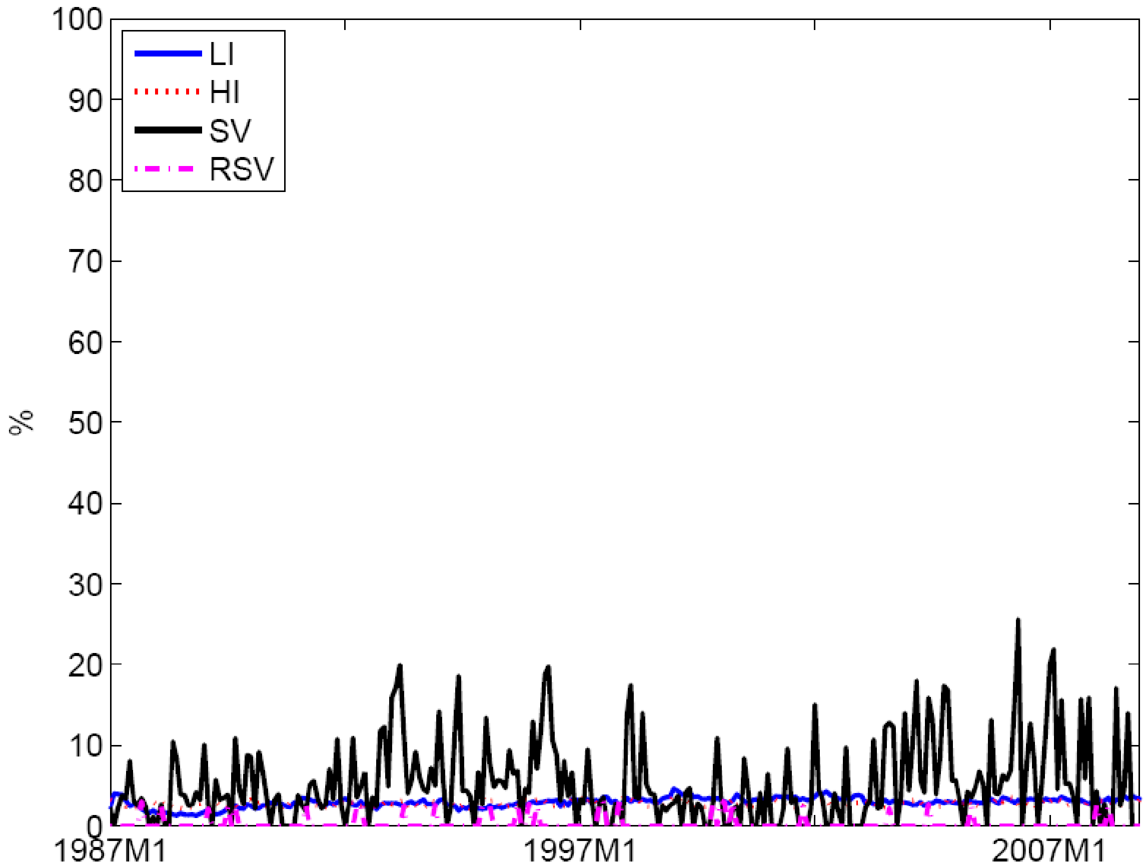
Economic gains: (transaction costs = 0.1%)

	model				combination scheme			
	1	2	3	4	1	2	3	4
	LI	HI	SV	RSV	BMA	LIN	TVW	RTVW
mean return	4.708	4.740	4.811	4.657	4.700	5.176	5.020	5.784
st dev return	0.794	0.769	1.139	0.614	0.739	4.355	1.332	3.062
Sharpe ratio	0.110	0.156	0.167	0.060	0.108	0.128	0.300	0.380
realized utility	-51.77	-51.77	-51.77	-51.79	-51.78	-51.75	-51.71	-51.58
Δ_s	284.7	287.9	284.5	276.6	279.1	279.1	311.7	373.6
Δ_m	-64.65	-61.42	-64.80	-72.72	-70.18	-52.18	-37.66	24.31
Δ_b	10.81	14.04	10.67	2.741	5.289	23.29	37.81	99.77

$\Rightarrow$ RTVW remains the best, keeping $\Delta_m > 0$, when transaction costs are taken into account.

Financial application: forecasting S&P 500 returns (continued)

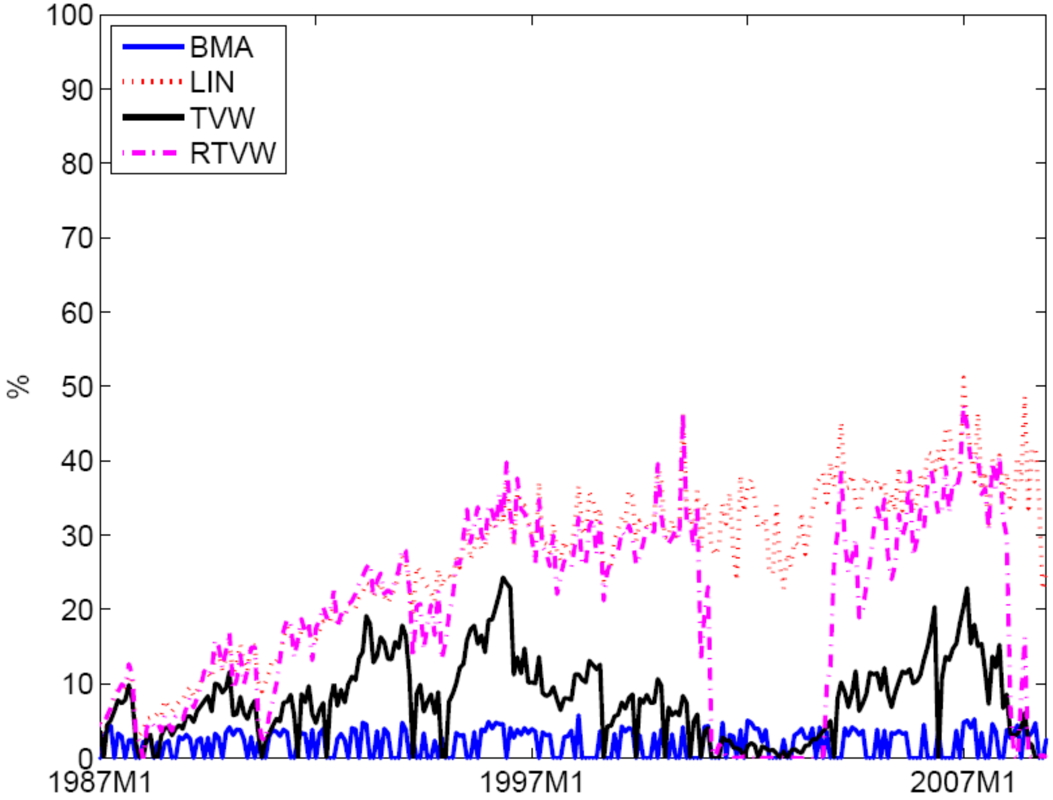
Figure: portfolio weight pw_{T+1} on risky asset (S&P500) in out-of-sample period for individual models (LI, HI, SV, RSV):



$\Rightarrow$ Individual models allocate too low weight pw_{T+1} to risky asset, resulting in low portfolio returns.

Financial application: forecasting S&P 500 returns (continued)

Figure: portfolio weight pw_{T+1} on risky asset (S&P500) in out-of-sample period for forecast combination schemes (BMA, LIN, TVW, RTVW):



- BMA allocates too low weight pw_{T+1} to risky asset (⇒ low portfolio returns)

- LIN, TVW,RTVW combinations allocate higher weights pw_{T+1} to stock asset.

- RTVW is the only scheme that drastically reduces this weight in bear market periods (burst of internet bubble in 2001-2003, recent financial crisis in 2nd part of 2007 and 2008).

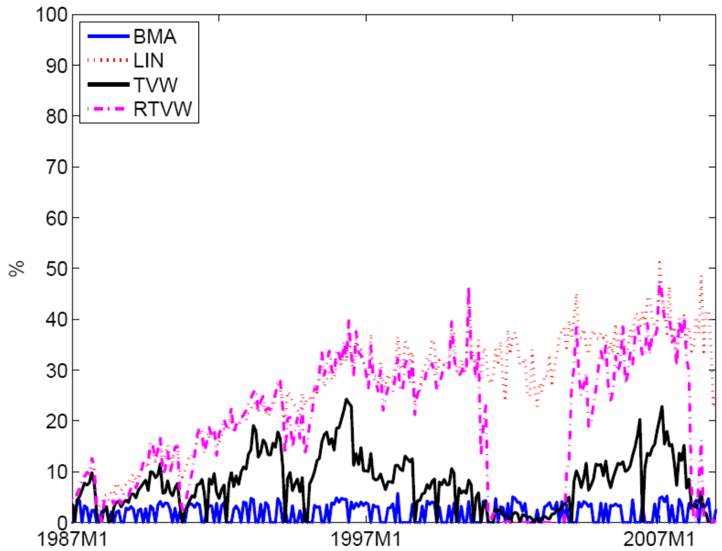
Financial application: forecasting S&P 500 returns (continued)

Robust, flexible structure of RTVW pays off:

- RTVW reduces weight pw_{T+1} in bear markets (compared with LIN, TVW)

- RTVW has higher weight pw_{T+1} in bull markets (compared with individual models and BMA). Reason: *'shrunk' predictive density*:

Figure: pw_{T+1} over time:



The 'shrunk' *excess* return distribution is **not** so much 'compressed' that pw_{T+1} switches from 0% to 100% when its mean changes from negative to positive values. (This behavior would result if the 'shrunk' density's st.dev would $\rightarrow$ 0.)

Rather, the parameter and model uncertainty incorporated in the 'shrunk' predictive density imply an investment strategy with a smooth, 'moderate', yet flexible evolvement over time for pw_{T+1}.

Financial application: forecasting S&P 500 returns (continued)

Lettau & Van Nieuwerburgh (2008, *Review of Financial Studies,* *"*Reconciling the return predictability evidence*"*):

> The uncertainty on the size of *steady-state shifts* rather than their dates is responsible for the difficulty of forecasting stock returns in real time.

The *'shrunk' predictive density* of the RTVW scheme may be particularly informative on the current and future evolvement of this *steady-state*, the driving force of return predictability.

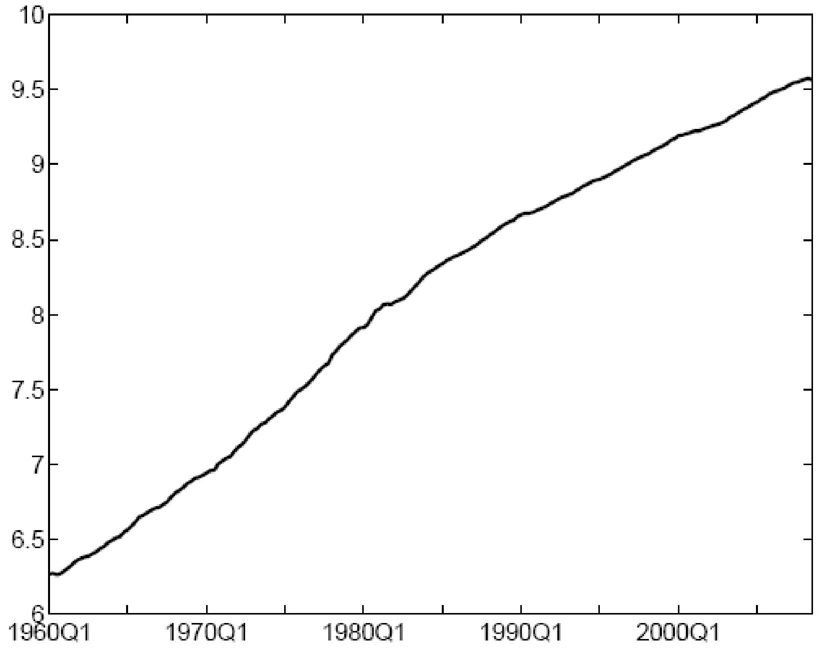
This may be the explanation for the RTVW scheme's good results.

We intend to analyze the RTVW scheme's performance in other portfolio management exercises in future research, to investigate the robustness of our findings.

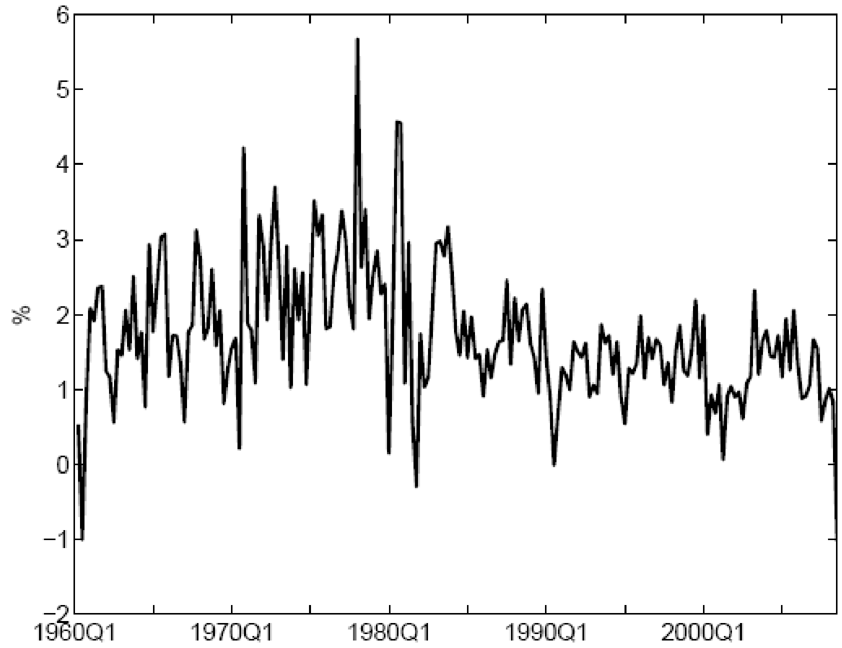
Macro application: forecasting US real GDP growth

Data: quarterly US real GDP growth (in %)

Period: in-sample: 1960:Q1 - 1979:Q4
out-of-sample: 1980:Q1 – 2008:Q3 (115 obs.)



Quarterly *log* levels of US real GDP



Quarterly US GDP growth rate (in %)
(= 100 x log difference)

Macro application: forecasting US real GDP growth (continued)

We use n = 6 individual models:

Model 1: Random Walk model (RW) *

Model 2: Random Walk model with drift (RWD) *

Model 3: AR(1) model. We follow Schotman & Van Dijk (1991, *J of Econometrics*, "A Bayesian Analysis of the Unit Root in Real Exchange Rates"), specifying a weakly informative *'regularization' prior* that helps to prevent problems that could be encountered during the estimation using the Gibbs sampler, if a flat prior were used.

Model 4: Error Correction Model (ECM) *, from De Pooter, Ravazzolo, Segers & Van Dijk (2008, *Advances in Econometrics*, "Bayesian Near-Boundary Analysis in Basic Macroeconomic Time-Series Models")

Models 5 & 6: the State-Space Model (SSM) and its robust extension (RSSM), given by the SV and RSV models of the financial application.

---

* Models 1, 2, 4: for log US real GDP (instead of US real GDP *growth*)

Macro application: forecasting US real GDP growth (continued)

For the ECM

$$\Delta y_t = \delta + (\rho_1 + \rho_2 - 1)(y_{t-1} - \mu - \delta(t-1)) - \rho_2(\Delta y_{t-1} - \delta) + \varepsilon_t$$

that can be rewritten as

$$y_t - \delta\, t = (1 - \rho_1 - \rho_2)\,\mu + \rho_1(y_{t-1} - \delta(t-1)) + \rho_2(y_{t-2} - \delta(t-2)) + \varepsilon_t$$

with $\varepsilon_t \sim N(0, \sigma^2)$, we specify a *'regularization' prior* that is an extension of Schotman & Van Dijk (1991).

Macro application: forecasting US real GDP growth (continued)

Table: Forecasting US real GDP growth (in %) : RMSPE

individual models		combination schemes	
1. RW	1.650	1. BMA	0.718
2. RWD	0.863	2. LIN	0.829
3. AR	0.772	3. TVW	0.757
4. ECM	0.790	4. RTVW	0.727
5. SSM	0.730		
6. RSSM	0.747		

$\Rightarrow$ - Random walk models (for log US real GDP) perform poorly.

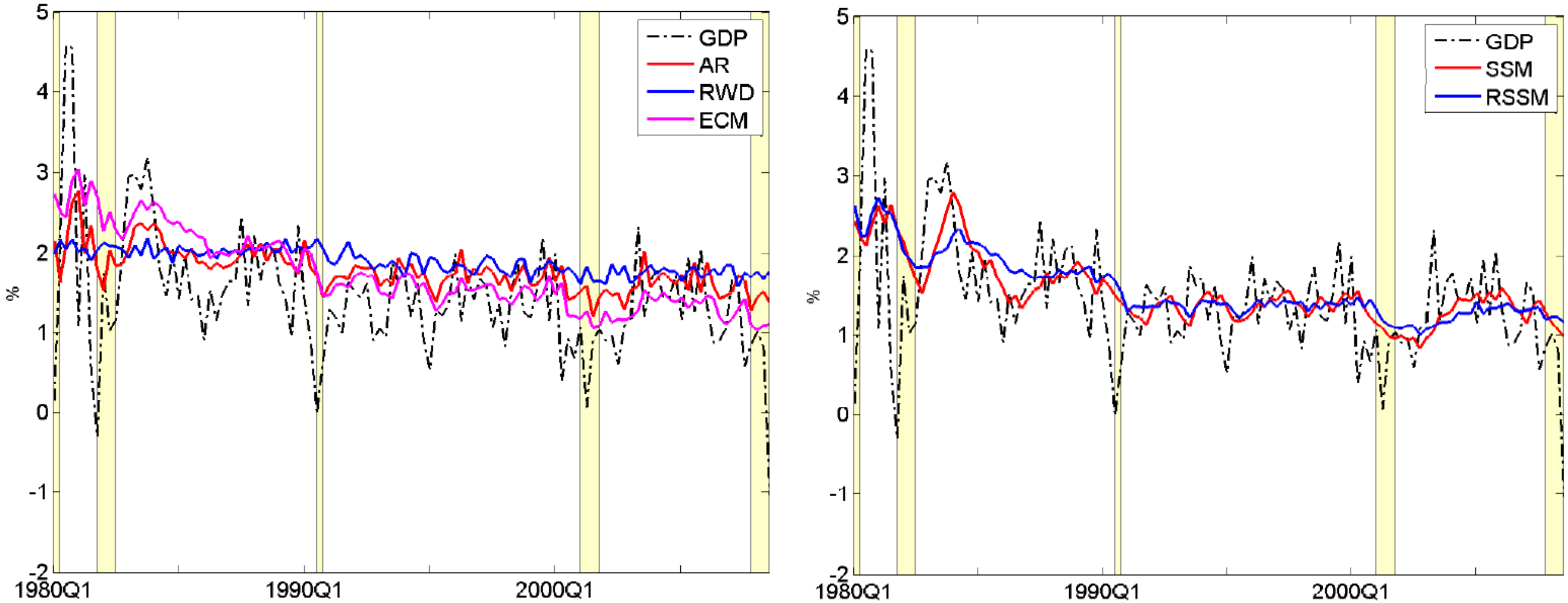
- For all other models, the test of Clark & West (2007, *J of Econometrics*) for equal forecasting quality of nested models rejects null versus RW.

- The models with time varying parameters, SSM and RSSM, perform well.

- BMA, RTVW combination schemes are even better than SSM, RSSM. LIN combination scheme performs poorly.

Macro application: forecasting US real GDP growth (continued)

Figure: Quarterly US real GDP growth (in %), point forecasts given by individual models. (Vertical bars highlight NBER recession periods.)

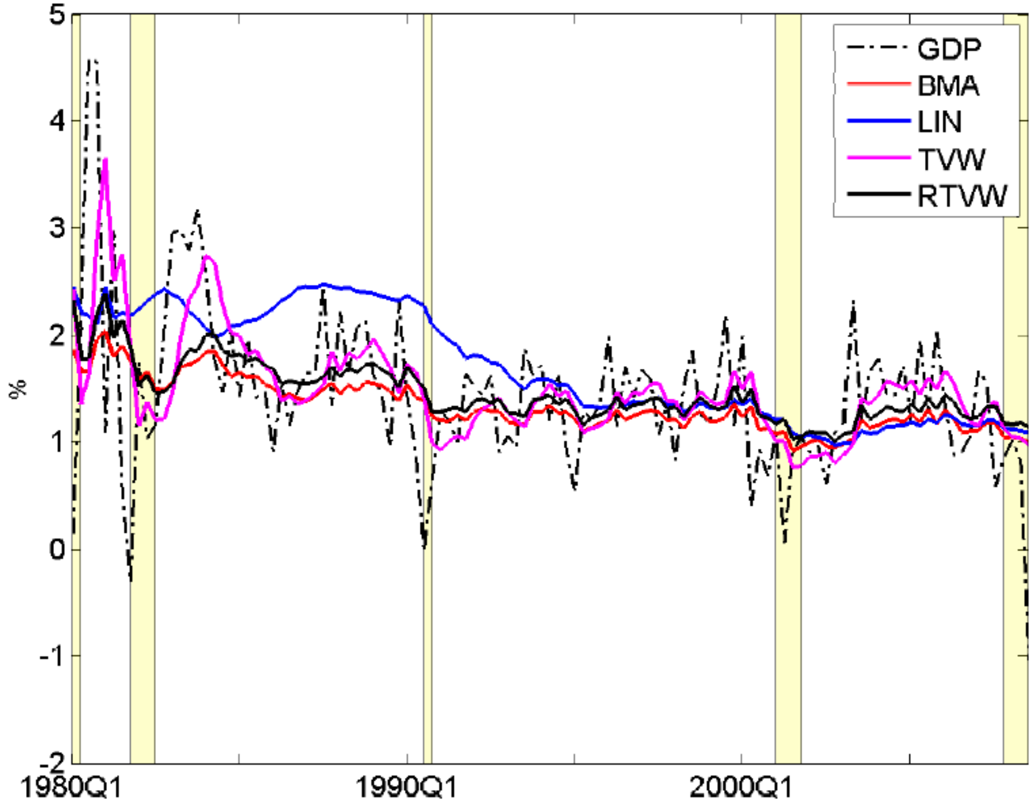


⇒ The models with fixed parameters (AR, RWD, ECM) perform poorly when GDP growth decreases rapidly as in NBER recessions. It takes some quarters for these models to adjust, in particular in 2001 and 2008 recessions.

The models with time-varying parameters (SSM, RSSM) cope better with this.

Macro application: forecasting US real GDP growth (continued)

Figure: Quarterly US real GDP growth (in %), point forecasts given by combination schemes. (Vertical bars highlight NBER recession periods.)



- LIN performs particularly poorly in 1980's & 1990's. Weight estimates for LIN may be inaccurate, as number of individual models $n = 6$ is relatively large and instability possibly high.

- BMA, TVW, RTVW react much faster to sharp decreases in GDP. Especially RTVW may early indicate recessions: before both 1991 & 2001 crises its point forecast decreases substantially with approximately 0.5%.

Final remarks

Findings in empirical applications:

- Forecast combination strategies can give higher predictive quality than selecting the best model;
- Properly specified time varying model weights yield higher forecast accuracy & economic gains compared with other schemes

Multiple directions for future research:

- a rigorous analysis of the impact of some assumptions (e.g. σ^2, Σ)
- a study on the robustness of the findings (e.g. for other data sets).
- comparison with other time varying weight combination schemes, e.g. regime switching (Guidolin and Timmermann (2007)), or schemes that carefully model breaks (Ravazzolo, Paap, Van Dijk & Franses (2007)).
- prediction of multivariate returns processes.
- specific prediction of variance, skewness or kurtosis (rather than mean).