RUHR ECONOMIC PAPERS

Dongkoo Kim
Tae-hwan Rhee
Keunkwan Ryu
Changmock Shin

Crowdsourcing of Economic Forecast – Combination of Forecasts using Bayesian Model Averaging

rwi #546

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #546

Responsible Editor: Roland Döhrn

ISSN 1864-4872 (online) – ISBN 978-3-86788-624-6

Ruhr Economic Papers #546

Dongkoo Kim, Tae-hwan Rhee, Keunkwan Ryu,
and Changmock Shin

Crowdsourcing of Economic Forecast – Combination of Forecasts using Bayesian Model Averaging

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788624
ISSN 1864-4872 (online)
ISBN 978-3-86788-624-6

Dongkoo Kim, Tae-hwan Rhee, Keunkwan Ryu,
and Changmock Shin[1]

Crowdsourcing of Economic Forecast – Combination of Forecasts using Bayesian Model Averaging

Abstract

Economic forecasts are quite essential in our daily lives, which is why many research institutions periodically make and publish forecasts of main economic indicators. We ask (1) whether we can consistently have a better prediction when we combine multiple forecasts of the same variable and (2) if we can, what will be the optimal method of combination. We linearly combine multiple linear combinations of existing forecasts to form a new forecast ("combination of combinations"), and the weights are given by Bayesian model averaging. In the case of forecasts on Germany's real GDP growth rate, this new forecast dominates any single forecast in terms of root-mean-square prediction errors.

JEL Classification: E32, E37

Keywords: Combination of forecasts; Bayesian model averaging

March 2015

1 Dongkoo Kim, Samsung Economic Research Institute (SERI); Tae-hwan Rhee, SERI and RWI; Keunkwan Ryu, Seoul National University; Changmock Shin, SERI *– We thank Roland Döhrn and the participants of the RWI Brown Bag seminar for valuable comments and suggestions.* *– All correspondence to: Tae-hwan Rhee, Research Fellow at SERI, Korea, e-mail: thrhee@alumni.stanford.edu*

1 Introduction

Economic forecast has never been easy. It's a task to predict the future value of an economic variable of interest, such as GDP growth rate, consumer price index, balance of international payments. It requires professional knowledge, skill and experience, and it depends on a vast amount of data. Considering that economics is a social science and that an economic variable is determined through almost infinite number of interactions among billions of human beings, it is close to impossible to make a correct and precise forecast. The famous assumption "ceteris paribus" will be almost always proven wrong.

Still, economic forecasts are quite essential in our daily lives. Businesses make decisions on production, investment, and labor compensation depending on the forecasts of market demand, business cycles, exchange rates, etc. for the next months or years. Households make consumption choices depending on the forecasts of income and consumer price movements, etc. In human capital investment decision, the forecast of each industry's growth rate is crucial, where the industry of concern may or may not even currently exist, and the time span easily goes beyond a couple of decades.

Because of this importance, many research institutions periodically make and publish forecasts of main economic indicators. Among these institutions are the international organizations such as the International Monetary Funds (IMF), the Organization for Economic Cooperation and Development (OECD), and the European Union (EU). In Germany, there are government-related or public institutions such as the *Rheinisch-Westfälisches Institut für Wirtschaftsforschung* (RWI), the German Institute for Economic Research (*Deutsches Institut für Wirtschaftsforschung*; DIW), and also private institutions such as the Cologne Institute for Economic Research (*Institut der deutschen Wirtschaft Köln*; IW). Investment banks such as Morgan Stanley and Citi Group also issue periodic economic forecasts, with regular or irregular intervals.

To make a forecast, all these institutions introduce some specific models and assumptions of their own regarding the future behavior of economic agents (household, business, and government). The differences in the models and assumptions will lead to the differences in the forecasts they make. For this reason, we typically have multiple forecasted values at hand for the same target variable for the same period, such as Germany's GDP growth rate in 2014.

This is the starting point of this paper. The central questions of this paper are (1) whether we can consistently have a better prediction when we combine multiple forecasts of the same target variable, and (2) if we can, what method of combination will be the best. Combination of forecasts basically means using more information. However, there is a good chance of reducing, rather than enhancing, predictive accuracy if we blindly add noisy or bad information to good information. Therefore it is essential to find an appropriate system of weighting each institution's forecast.

In this paper, we introduce a way of applying the method of Bayesian model averaging to the question of forecast combination. We combine different linear combinations of existing forecasts to form a new forecast ("combination of combinations"), and the weights are given by Bayesian model averaging. When we apply this method to the forecasts of Germany's GDP growth rates, combining the forecasts made by 6 different institutions, the new forecast is shown to produce a more accurate out-of-sample prediction than the original forecasts. It dominates any single forecast in terms of root-mean-square prediction errors (RMSPE).

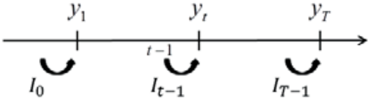
Here is what follows. In Section 2 we introduce the model and the method of Bayesian model averaging. Section 3 deals with the data used in this paper. Section 4 shows the application of our methodology using Germany's GDP forecasts made by 6 different institutions and summarizes the results. Section 5 concludes with implications and direction of further studies.

2 Model

2.1. Outline of the model

Let y be the variable of interest, such as real GDP growth rate. There are k different forecasts of y, denoted by $x_1, x_2, \cdots, x_k$. For now, we only deal with one-period ahead forecasts of y, but it can be easily extended. Let I_{t-1} be the information set available at time $t-1$. Denoting institution j's forecast of y_t at time $t-1$ by $x_{t,j}$,

$$\begin{pmatrix} y_{t-1} \\ x_{t,1} \\ x_{t,2} \\ \vdots \\ x_{t,k} \end{pmatrix} \in I_{t-1} .$$



Note that for each t, $y_{t-1}, x_{t,1}, \cdots, x_{t,k}$ are in the information set I_{t-1}, but not $y_t \in I_{t-1}$.

Now consider constructing a model to form a new forecast of y. Here, a model corresponds to a so-called "combination of forecasts." Assume that we are at T, and want to forecast y_{T+1}. Then the model is

$$y_{T+1} = \beta_0 + \beta_1 x_{T+1,1} + \beta_2 x_{T+1,2} + \cdots + \beta_k x_{T+1,k} + e_{T+1} . \quad (1)$$

The relevant information set is I_T, i.e. $y_1, y_2, \cdots, y_T$ are known, and we know the past forecasts $x_{1,j}, x_{2,j}, \cdots, x_{T+1,j}$ for each institution $j \in \{1,2,\cdots,k\}$. In order to estimate $\beta_0, \beta_1, \cdots, \beta_k$ in the final model, we first construct and estimate "interim" models in the same formulation as (1), then we combine these interim models using Bayesian model averaging method. Hence we call our method "combination of combinations."

Each interim model can contain a different number of forecasts, from 1 to k. If we order the forecasts in advance, we may consider k different interim models, the first containing only the best forecast, the second being a combination of two best forecasts, etc. Alternatively, we may have at most $2^k - 1$ different models à la Sala-i-Martin et al (2004), where each forecast may or may not be used in a linear combination of forecasts, and at least one forecast needs to be included in any given combination. Following the first method of choosing models with pre-ordering, we consider k different interim models: $C_1, C_2, \cdots, C_k$.

Several methods of ordering the existing forecasts can be employed, namely:

(a) RMSE(root-mean-square error) or MAPE(mean absolute percentage error)
(b) sequential/stepwise R^2 criteria (as in Liang and Ryu. preferred), or
(c) subjective judgement, etc.

When we apply this method to Germany's growth rate forecasts in Section 4, we will use the sequential R^2 criteria. To save on notations, assume that $x_1 \succ x_2 \succ \cdots \succ x_k$ in terms of prediction accuracy, where $A \succ B$ means *A* is preferred to *B*. Then the interim Model j (C_j) uses only $x_1, x_2, \cdots, x_j$, and it has $j + 1$ parameters to estimate including constant.[2] For generality purpose, let us denote the number of parameters in Model j as k_j.

$$C_j: y = \beta_0 l + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + e$$

where y is a $T \times 1$ vector of realized values, l is a $T \times 1$ vector of ones, and x_j is a $T \times 1$ vector of jth institution's forecasts.

[2] *Granger and Ramanathan(1984) showed that, when making a linear combination of forecasts, the best method is to add a constant term and not to constrain the weights to add to unity.*

$$\Leftrightarrow C_j: y = X_j \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_j \end{pmatrix} + \mathrm{e}$$

where $X_j = (l \,\vdots\, x_1 \,\vdots\, x_2 \,\vdots\, \cdots \,\vdots\, x_j)$ is a $T \times (j+1)$ matrix.

$$\Leftrightarrow C_j: y = X\beta^j + e$$

where $X = X_k$ and $\beta^j \equiv \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_j \\ \cdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}$.

Note that the last $k - j$ elements in the $(k+1) \times 1$ vector β^j are restricted to be equal to zero, i.e. $\beta_{j+1} = \cdots = \beta_k = 0$.

2.2. Bayesian posterior on $\boldsymbol{\beta}$

To get $\hat{\beta}$, the Bayesian estimator of β in the final model, we first formulate a Bayesian posterior density function of β conditional on data y. Following the method of Bayesian model averaging, this Bayesian posterior can be written as[3]

$$g(\beta|y) = \sum_{j=1}^{k} P(C_j|y) \left[\frac{g(\beta|C_j) f(y|C_j, \beta)}{f(y|C_j)} \right]$$

which is weighted average of model-specific posteriors with weight being equal to $P(C_j|y)$, posterior model probabilities.

Note that, under non-informative (diffuse) priors, each model-specific posterior

[3] *See, for example, Zellner(1971), Leamer(1978), Sala-i-Martin et al.(2004), or Hansen(2007).*

$$\left[\frac{g(\beta|C_j)f(y|C_j,\beta)}{f(y|C_j)}\right]$$

is nothing else but the sampling distribution of OLS estimator $\hat{\beta}^j$ from the Model j

$$C_j: y = X\beta^j + e$$

which is distributed according to $N\left(\hat{\beta}^j, Var(\hat{\beta}^j)\right)$

where $\hat{\beta}^j = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_j \\ \cdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, $Var(\hat{\beta}^j) = \begin{pmatrix} Var\begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_j \end{pmatrix} & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{pmatrix}$.

2.3. Bayesian posterior model probabilities, $P(C_j|y)$

As the model-specific prior information becomes dominated by the data information for each model, i.e. $X_j{}'X_j \to \infty$ for each $j \in \{1,2,\cdots,k\}$, we have, for C_i and C_j with $i \neq j$,

$$\frac{P(C_i|y)}{P(C_j|y)} \approx \frac{P(C_i)T^{-k_i/2}SSE_i{}^{-T/2}}{P(C_j)T^{-k_j/2}SSE_j{}^{-T/2}} \quad \text{.........} \ (2)$$

where T = number of observations used to estimate the models

$P(C_j)$ = prior model probability

k_j = number of parameters in Model j

SSE_j = sum of squared errors from OLS estimation of Model j.

As is clear from the above formula (2), the posterior odd ratio between two models, C_i and C_j, is a product of the following three terms:

(i) prior odd: $P(C_i)/P(C_j)$

(ii) penalty for lack of "parsimoniousness": $T^{-k_i/2}/T^{-k_j/2}$

(iii) penalty for lack of "in-sample performance":

$$SSE_i^{-T/2}/SSE_j^{-T/2}$$

From the posterior odd ratio (2), we derive the posterior model probability up to a proportionality constant as

$$P(C_j|y) \propto P(C_j)T^{-k_j/2}SSE_j^{-T/2}$$

$$\log P(C_j|j) = const + \log P(C_j) - \frac{1}{2}[k_j \log T + T \log SSE_j] \;.$$

It is no wonder that the expression in square brackets is the Bayesian Information Criterion (BIC_j), which is often used for model selection purpose.

$$k_j \log T + T \log SSE_j = BIC_j \qquad \text{...............} \; (3)$$
$$\Rightarrow \exp(BIC_j) = T^{k_j} SSE_j^T$$
$$\Rightarrow P(C_j|y) \propto P(C_j)[\exp(BIC_j)]^{-1/2}$$
$$= P(C_j)\exp[-\frac{1}{2}BIC_j]$$

In summary, the posterior model probability is determined by

(i) prior model probability $P(C_j)$
and (ii) model selection criteria BIC_j

To calculate the posterior model probability, all we have to do now is to specify the model prior $P(C_j)$ in some proper way. Let's consider a generalized weighting scheme

$$P(C_j) \propto 1 + \omega + \cdots + \omega^{j-1}$$

using a real number $\omega \in [0,1]$. In the following sections we will try two values, $\omega = 0$ and $\omega = 0.5$, and compare the results to each other. Note that the $\omega = 0$ case corresponds to an equal weighting scheme

$$P(C_1) = P(C_2) = \cdots = P(C_k) = 1/k \;.$$

This specification gives equal prior probability to each model, not to each forecast.[4] Then, in the case of $\omega = 0.5$, we have

$$P(C_j) = const \cdot [1 + \tfrac{1}{2} + \cdots + \left(\tfrac{1}{2}\right)^{j-1}]$$

which assigns higher prior probability to models containing larger number of forecasts.[5]

Once we specify the model prior $P(C_j)$, we have

$$P(C_j|y) = \frac{P(C_j)[\exp(BIC_j)]^{-1/2}}{\sum_{i=1}^{k} P(C_i)[\exp(BIC_i)]^{-1/2}} \quad \text{.......... (4)}$$

2.4. Inferences on $\boldsymbol{\beta}$ based on $\boldsymbol{g(\beta|y)}$

We have shown that (i) the Bayesian posterior density function $g(\beta|y)$ can be written as the expectation of model-specific posteriors weighted by posterior model probabilities, and (ii) the model posteriors are determined by model priors and model selection criteria BIC_j.[6]

After we estimate each one of k different models, $C_1, \cdots, C_k$, by OLS, we can derive the expectation of β using its posterior density as

$$E(\beta|y) = \sum_{j=1}^{k} P(C_j|y)\, \hat{\beta}^j \qquad \text{.................... (5)}$$

---

[4] *Among the k different forecasts, x_1 is contained in all of the k models, but x_k is used only in the kth model C_k. If we assign equal prior probability to every model, we're putting the largest weight on x_1 and the smallest weight on x_k.*

[5] *Assigning higher prior to models with more forecasts does not mean that each forecast has equal weight. The number of forecasts contained in C_2 is twice as large as that of C_1, but the prior on C_2 is only 1.5 times larger than that on C_1. So we're putting larger weight on the models using more forecasts, but not in strict proportion to the number of forecasts contained in each model.*

[6] *One may use alternative model selection criteria for BIC_j, such as AIC (Akaike Information Criterion) or an increasing transformation of Mallows' C_P :*

(i) $BIC_j = k_j \log T + T \log SSE_j$

(ii) $AIC_j = 2k_j + T \log SSE_j$

(iii) $T \log C_P$, where $C_P = 2\sigma^2 k_j + SSE_j$ is the so-called Mallows' C_P.

where the $(k+1) \times 1$ vector

$$\hat{\beta}^j = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_j \\ \cdots \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ is given by OLS estimation of } C_j,$$

and is equal to $E(\beta|C_j, y)$.

Using the variance decomposition formula[7],

$$Var(\beta|y) = E_{C_j} Var(\beta|C_j, y) + Var_{C_j}(E[\beta|C_j, y]) \quad \text{..........} (6)$$

$$= \sum_j P(C_j|y) Var(\beta|C_j, y)$$

$$+ \sum_j P(C_j|y)(\hat{\beta}^j - E(\beta|y))(\hat{\beta}^j - E(\beta|y))'$$

where the $(k+1) \times (k+1)$ matrix

$$Var(\beta|C_j, y) = \begin{bmatrix} Var\begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_j \end{pmatrix} & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix}.$$

Here $Var\begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_j \end{pmatrix}$ for each Model j is available from OLS estimation on C_j, namely

$$Var(\beta|C_j, y) = \begin{bmatrix} (X_j' X_j)^{-1} \hat{\sigma}_j^2 & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix}.$$

[7] $Var(\cdot) = E$("within-model variance") + "between-model variance"

Using the formula in (6), we can compute a posterior 95% probability interval for β_j as $\sum_{i=1}^{k} P(C_i|y) \cdot [\hat{\beta}^i]_j \pm 2\, SE(\beta_j|y)$. [8]

3 Data

Here we apply the methodology introduced in the previous section to a real-world economic forecast. The target variable is Germany's real GDP growth rate, in percentage term. For example it was 0.4 for 2013 and 3.6 for 2010. To build a correct series of this variable, we need to keep in mind that there were multiple occasions of changes in Germany's GDP accounting method. For example, it was revised in the 2nd quarter of 1999 to comply with the new set of guidelines in ESA 1995 [9]. Other major updates include the transition to chain-index pricing in 2005 and the introduction of new industrial classification system in 2011. We also need to consider that realized GDP growth rates themselves are updated over time. In this paper, we use the first release data of the quarterly national account which are published 50 days after the end of target years.

We use forecasts by 6 different institutions. Three of them are domestic: the *Rheinisch-Westfälisches Institut für Wirtschaftsforschung* (RWI), the Joint Forecast (*Gemeinschaftsdiagnose*; GD), and the German Council of Economic Experts (*Sachverständigenrat*; SVR). The other three are international organizations: the International Monetary Fund (IMF), the Organization for Economic Cooperation and Development (OECD), and the European Union (EU). The time span is from 1991 to 2013. Since we use yearly forecasts, we have 23 observations.

---

[8] *For the individual coefficient on jth forecast, β_j,*

$$SE(\beta_j|y) = \sqrt{Var(\beta|y)_{(j+1)(j+1)}}\ ,$$

which is the square root of the $(j+1)$st diagonal element of $Var(\beta|y)$ in (6).

[9] *European System of National Accounts (Europäische System Volkswirtschaftlicher Gesamtrechnungen; EVSG) 1995, which is the European version of United Nation's SNA (System of National Accounts) 1993.*

We also need to pay detailed attention to the timing of the forecasts. Five out of the six institutions publish their growth forecasts at least twice a year. Thus, we need to choose among multiple forecasts for each institution. In this paper, for the forecast of GDP growth in year t, we use the latest forecast available at the end of December in year $t-1$ for each institution. For example, for Germany's real GDP growth rate forecast for year 2013, we take each institution's latest forecast available on December 31, 2012, and denote these forecasts as $x_{2013,RWI}, x_{2013,GD}, x_{2013,SVR}, \cdots$, etc.

Table 1 summarizes the descriptive statistics of Germany's real GDP growth rate forecasts from the 6 institutions, and the realized values. This table shows a couple of interesting characteristics. First, for every one of the 6 institutions, the mean of forecasts is higher than the mean of actual growth rates. In other words, all six institutions have made more optimistic forecasts on average than the actual status of German economy since 1991. For RWI, SVR, EU, and OECD, the gap is 0.3~0.4%p. Forecasts of GD and IMF are notably optimistic: the means of their forecasts are higher than that of the realized values by 0.44 and 0.58%p, respectively.

Second, the standard deviation of forecasts is significantly smaller than that of actual growth rates, again for every one of the 6 institutions. The standard deviation of realized values is 1.85, while the standard deviations of all 6 forecasts are gathered around 1. This means that only about half ~ two-thirds of the movements of actual growth rates are reflected in the forecast series. To put it differently, the forecasts of these 6 institutions can be characterized as "conservative."

This conservatism is easily revealed when one compares Figure 1 and Figure 2. When the actual growth rates are higher than average, the forecasts have a strong tendency to be lower than the realized values, and vice versa.

Table 2 shows the bilateral correlation coefficients among the forecasts of Germany's real GDP growth rates and the realized values. Note that the correlation coefficients are all positive among the 6 institutions — in general they have moved in the same direction. The correlation between SVR and EU is particularly high at 0.970, while that between SVR and IMF is relatively lower at 0.711. These 6 forecasts have been moving in similar directions during this 23-year period, but the specific direction and magnitude of movements vary year by year.

There is wider variation among the bilateral correlation coefficients between the forecasts and the realization, than among those between the forecasts. RWI's forecast shows a rather high correlation, 0.815, with the actual growth rate, while IMF shows a lot lower correlation, 0.319.

In summary, considering the RMSE and correlations between the forecasts and realization, RWI's forecast has been the most precise, with the smallest RMSE and the highest correlation coefficient. IMF's forecast by itself can be evaluated as the least accurate, with the largest RMSE and the lowest correlation.[10]

4 Results

Using the methodology described in Section 2 and Germany's GDP growth rate forecasts summarized in Section 3, now we make a new "combination of combinations" forecast based on the forecasts of these 6 institutions. Our goal is to show that, if we make a linear combination of the forecasts and assign proper weight to each forecast according to the method of Bayesian model averaging, we will be able to form a new, more informative forecast. The first step of this procedure is a set of stepwise regressions, through which we can order the forecasts according to

[10] *Of course we have to take into account that there are differences in the timing of forecasts. The IMF's forecasts are published in Septembers, while the RWI's forecasts are published in Decembers. This gives the RWI a significant advantage in terms of accuracy. However, the IMF's early forecast, or the difference between the IMF's forecast and the RWI's forecast, still contains important information, as you will see in the next section.*

the in-sample fitting performance: from the one most fitted to the realized values, to the one least informative.

In the first round, we run 6 regressions in total. For each regression we try each one of the 6 forecasts as the explanatory variable, while the realized value of Germany's GDP growth rate is the dependent variable. The coefficient of determination (R^2) is the highest when RWI's forecast is used. In other words, RWI's forecast has the largest explanatory power in the case of 1-variable regression.[11] Let us denote these 1-variable regression models as Model a, b, ⋯, f, respectively. The results of the first round regressions are summarized in Table 3.

Now that RWI's forecast has been found to be the best fit to the realized values, we use this forecast as a fixed explanatory variable in all of the second round regressions. In this round we run 2-variable regressions, five of them in total. For each regression RWI’s forecast and one of the other 5 institutes' forecasts are used as explanatory variables. Now we have the largest R^2 when IMF’s forecast is added. (Table 4) Note the change in the role of IMF’s forecast between the first and the second round regressions. In the 1-variable regression, the explanatory power of the IMF's forecast is very low, with the coefficient of determination being just 0.102. However, working as an additional variable given RWI's forecast, it turns out that IMF's forecast is the most informative in predicting the next year’s growth rate (biggest marginal contribution to predictability).

Now fix RWI and IMF's forecasts as explanatory variables, and put a forecast from one of the 4 remaining forecasts as the third explanatory variable. In this round with four different 3-variable regressions, pick the one which leads to the largest coefficient of determination. Go on to the fourth round with 4-variable regressions, then to the final round with 5-variable regressions. In this way we re-order all 6 institutes ranked according to the

---

[11] As can be seen in Table 3, the actual number of regressors is 2 including the constant rather than 1. We'll still call this round as "1-variable regressions", emphasizing that "1" is the number of forecasts being used in those regression equations.

(additional) predictive powers of their forecasts: RWI, IMF, EU, SVR, OECD, and GD.

As a result of the previous stepwise regressions, we now have six($k = 6$) interim models, each one being a combination of forecasts. The first model includes only RWI's forecast as an explanatory variable. From here on we call RWI as Institute 1, and denote this model as C_1, reading "combination one." IMF becomes Institute 2, and the interim model with Institutes 1 and 2 is C_2, etc. The regression results of $C_1, C_2, \cdots, C_6$ are summarized in Table 5.

Now we evaluate the Bayesian posterior model probability for each one of these 6 interim models. For this we need to specify the prior probabilities. As is mentioned in Section 2, there are many ways to set the model prior. Here we consider two cases: the equal weighting scheme ($\omega = 0$) and assigning higher prior probability to a model with more explanatory variables($\omega = 0.5$). With a set of properly specified model priors, the posterior model probability $P(C_j|y)$ can be derived through Formula (3) and (4). Table 6 compares the prior and posterior model probabilities for each interim model.

In both cases, with equal and increasing priors, the highest posterior model probability is assigned to Model 3(C_3). Except for Models 2 and 3, no other model is assigned a higher posterior probability than its prior, and Model 2's posterior probability is only slightly higher than its prior. When making forecasts on Germany's annual real GDP growth rates from 1991 to 2013, it is more probable for Model 3, which linearly combines three forecasts made by Institutes 1, 2, and 3, to be a better model than any other model with a different number of explanatory variables. And this result is not sensitive to the way we assign prior model probabilities.

Selecting Model 3 is not our final destination, of course. The last step is to combine the six models with the weights being the posterior model probabilities. This "combination of combinations" will be our final forecast model. Putting the model-by-model regression results summarized in Table 5 and the posterior probabilities in Table 6 into Formula (4), (5), and (6) gives us the final set of results, i.e. expected values and standard errors of the regression coefficients (forecast weights) according to the posterior distributions. These results are summarized in Table 7.

Table 7 shows two different sets of coefficients for the final model, according to the prior probabilities. The RMSE's of these final models are 0.617 and 0.616, which are quite lower than the RMSE 1.069 of Model 1(C_1). Model 1's explanatory variable is Institute 1's forecast, which has the largest explanatory power among the 1-variable regressions using only one institute's forecast as an explanatory variable. Therefore, our final model ("combination of combinations") shows a better performance than any single institute's forecast with regard to in-sample fitting.

Of course we have to pay attention to the number of explanatory variables when we try this sort of interpretation. In regression analyses, a higher number of explanatory variables mechanically leads to a smaller RMSE within sample period. So it is natural to expect our final model, which utilizes multiple forecasts, to have a smaller RMSE than any other model using only one forecast. To properly evaluate the in-sample performance of our model, therefore, we need to take into account the "effective" number of explanatory variables (ENEV) in our final model, which we will define as

$$\text{ENEV} \equiv \sum_{j=1}^{k} j \cdot P(C_j|y) \ .$$

Except for constants, Model 1(C_1) has one effective explanatory variable, i.e. ENEV = 1, and Model 2's ENEV is two. The final model is a weighted average of these interim models, which leads us to consider that the ENEV in the final model is the weighted average of ENEV's in the interim models. In the case of increasing priors($P(C_1) < P(C_2) < \cdots < P(C_6)$), the effective number of explanatory variables can be calculated as

$$2.961 = 0.0001 \times 1 + 0.1655 \times 2 + \cdots + 0.0012 \times 6 \ .$$

Now let us evaluate again the in-sample fitting performance of the final model in terms of RMSE, considering the notion of ENEV above. In the case of increasing priors, the ENEV is 2.961 and the RMSE is 0.616. This is smaller than 0.714, the RMSE of Model 2(C_2) with ENEV=2, and only slightly larger than 0.613, the RMSE of Model 3(C_3) with ENEV=3. This does not lead to an unambiguous conclusion that the final model's in-sample performance is better than any interim model, i.e. any linear combination without the process of Bayesian model averaging. However, considering that 2.961 is between 2 and 3, if we go from 0.714 toward 0.613 by 96.1%, we'll be at 0.617, which is only slightly larger than 0.616.

By nature of forecasting, out-of-sample performance is a more important criterion than in-sample performance. Our data period is not very long, which prevents us from meaningfully testing the fitness of out-of-sample predictions. Nevertheless, we try to see the out-of-sample prediction performance of our final model, using whatever data available to us.

Figure 3 shows the absolute values of forecast errors from year 2000 to 2013. Here the forecasts of our final model (C^2) are calculated recursively[12]. For example, for 2000's forecast, we derive our final model using the forecasts by 6 institutions and realization of GDP growth rates during 1991~1999 period, then we combine the 6 institutions' 2000 forecasts using the coefficients of the final model. The 2001~2013 forecasts of our final model are calculated in the same, recursive way. The 'increasing' prior model probabilities are assigned for this exercise.

Note that our final model beats any other forecast institution in 8 years, in 2001, 2003, 2005~7, and 2009~11, out of this 14-year period in terms of absolute forecast errors. 2001 and 2009~2011 were the years right after the dot-com bubble burst and the 2008~2009 Global Financial Crisis, respectively. Although these

[12] *Here our final model is denoted as C^2, signifying that it is a "combination of combinations".*

crises led to unusually inaccurate forecasts made by any institute, our final model’s out-of-sample performance can still be considered impressive. This result is summarized by Figure 4 from a different angle, which compares the root-mean-square prediction errors(RMSPE's) of the forecasts depicted in Figure 3 for the entire out-of-sample period, 2000~2013. Our final model has RMSPE of 1.07, which beats any single institution's forecast in that 14-year period.

5 Conclusion

Can we make a new forecast that is more precise, when we linearly combine multiple existing forecasts on the same target variable? This question was the starting point of our paper. We first showed that the method of Bayesian model averaging could be applied as the weighting scheme here. We constructed multiple linear models, and then we evaluated the posterior model probabilities of these interim models according to Bayesian theory. Our final model, which we called "combination of combinations", was the combination of the interim models using the posterior probabilities as the weights.

Against this theoretical background, we applied our method to forecasts of Germany’s GDP growth rates made by six different institutions. The final model we derived accordingly indeed beat any single forecast in terms of root-mean-square prediction errors, for the period of 2000~2013. Although the data length was not very long, we had a favorable signal that our method could actually be used to improve the precision of economic forecasts by combining multiple existing forecasts and/or multiple forecasting methods.

Our method has a wide range of application. This "combination of combinations" method can be applied in the same way to any field of interest in which we have multiple existing forecasts on a single target variable: current account balances, international oil prices, stock market indices, to name a few. Some of these forecasts have been more frequently released than yearly GDP growth rates, which leads to longer time series. In this case we expect that statistical significance of the coefficients in our final model will also be improved.

References:

Various Institutions, Periodical reports and papers on Germany′s economic forecast.

Granger, C. and R. Ramanathan (1984), Improved Methods of Combining Forecasts. Journal of Forecasting 3: 197-204.

Hansen, B. (2007), Least Squares Model Averaging. Econometrica 75: 1175-1189.

Leamer, E. (1978), Specification Searches: Ad Hoc Inference with Non-experimental Data. New York: John Wiley & Sons.

Leamer, E. (1983), Let's Take the Con Out of Econometrics. The American Economic Review 73: 31-43.

Liang, K. and K. Ryu (1996), Selecting the Form of Combining Regressions Based on Recursive Prediction Criteria. In J. Lee, W. Johnson, and A. Zellner (Eds.), Modelling and Prediction Honoring Seymour Geisser: 122-135. New York: Springer.

Liang, K. and K. Ryu (2003), Relationship of Forecast Encompassing to Composite Forecasts with Simulations and an Application. Seoul Journal of Economics 16: 363-386.

Sala-i-Martin, X., G. Doppelhofer and R. I. Miller (2004), Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. The American Economic Review 94: 813-835.

Zellner, A. (1971), An Introduction to Bayesian Inference in Econometrics. New York: John Wiley & Sons.

Zellner, A. (1986), On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions. In Prem Goel and Arnold Zellner (Eds.), Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti. Studies in Bayesian Econometrics and Statistics Series, Volume 6. Amsterdam: North-Holland.

Table 1: Descriptive Statistics, 1991-2013

Forecasts and realization of Germany's GDP growth rates in %

	SVR	RWI	EU	GD	IMF	OECD	Realized Values
Mean	1.64	1.68	1.65	1.76	1.90	1.70	1.32
Standard Deviation	0.95	1.20	1.01	0.78	0.92	1.00	1.85
RMSPE	1.50	1.17	1.54	1.57	1.87	1.43	

Note: RMSPE(root-mean-square prediction error) $= \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - g_t)^2}$*, where* x_t *is a forecast and* g_t *is the realized value of year* t*'s real GDP growth rate. In other words, RMSPE is the square-root of the average of squared forecast errors. The smaller an RMSPE is, the more precise a forecast is.*

Table 2: Correlation coefficients among growth forecasts and realization

	RWI	EU	GD	IMF	OECD	Realized Values
SVR	0.896	0.970	0.939	0.711	0.900	0.620
RWI		0.881	0.914	0.745	0.936	0.815
EU			0.957	0.733	0.905	0.578
GD				0.815	0.926	0.600
IMF					0.797	0.319
OECD						0.677

Note: Bilateral correlation coefficients among the forecasts of Germany's real GDP growth rates and the realized values in % terms.

Table 3: Stepwise regression, 1-variable cases

	Spec. a	**Spec. b**	Spec. c	Spec. d	Spec. e	Spec. f
dependent variable(y)	Germany's real GDP growth rate (yearly, %)					
regressors						
constant	-0.663 (0.632)	**-0.784** **(0.401)**	-0.411 (0.627)	-1.183 (0.796)	0.110 (0.873)	-0.810 (0.586)
SVR	1.208* (0.333)					
RWI		**1.253*** **(0.194)**				
EU			1.052* (0.324)			
GD				1.426* (0.414)		
IMF					0.638 (0.413)	
OECD						1.254* (0.298)
R^2	0.385	**0.665**	0.334	0.361	0.102	0.458

*Notes: standard errors in parentheses. * denotes p<0.05.*

Table 4: Stepwise regression, 2-variable cases

	Spec. ba	Spec. bc	Spec. bd	**Spec. be**	Spec. bf
dependent variable(y)	Germany's real GDP growth rate (yearly, %)				
regressors					
constant	-0.286 (0.438)	-0.330 (0.392)	0.823 (0.559)	**0.426** **(0.366)**	-0.275 (0.446)
SVR	-1.097* (0.514)				
RWI	2.029* (0.405)	2.103* (0.361)	2.485* (0.387)	**1.993*** **(0.199)**	2.253* (0.513)
EU		-1.143* (0.428)			
GD			-2.081* (0.597)		
IMF				**-1.291*** **(0.259)**	
OECD					-1.287* (0.618)
R^2	0.727	0.753	0.791	**0.851**	0.724

*Notes: standard errors in parentheses. * denotes p<0.05.*

Table 5: Stepwise regression results — interim combinations

	Model 1 (C_1)	Model 2 (C_2)	Model 3 (C_3)	Model 4 (C_4)	Model 5 (C_5)	Model 6 (C_6)
dependent variable(y)	Germany's real GDP growth rate (yearly, %)					
regressors						
constant	-0.784 (0.401)	0.426 (0.366)	0.598 (0.330)	0.660 (0.361)	0.645 (0.375)	0.693 (0.541)
RWI	1.253* (0.194)	1.993* (0.199)	2.490* (0.260)	2.542* (0.288)	2.482* (0.377)	2.498* (0.409)
IMF		-1.291* (0.259)	-1.142* (0.235)	-1.157* (0.242)	-1.181* (0.267)	-1.162* (0.313)
EU			-0.784* (0.303)	-0.539 (0.602)	-0.581 (0.639)	-0.529 (0.775)
SVR				-0.319 (0.675)	-0.316 (0.693)	-0.314 (0.714)
OECD					0.133 (0.517)	0.135 (0.533)
GD						-0.115 (0.912)
R^2	0.665	0.851	0.890	0.891	0.891	0.891
RMSE	1.069	0.714	0.613	0.610	0.609	0.608

*Notes: standard errors in parentheses. * denotes p<0.05. The order of regressors(forecasting institutes) are re-arranged according to the explanatory contribution.*

Table 6: Different priors and resulting posterior model probabilities

Model Probabilities		Model 1 (C_1)	Model 2 (C_2)	Model 3 (C_3)	Model 4 (C_4)	Model 5 (C_5)	Model 6 (C_6)
$P(C_j) = 1/6$	Prior	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667
	Posterior	0.0001	0.1893	0.7078	0.0913	0.0103	0.0011
$P(C_1) < \cdots < P(C_6)$	Prior	0.0997	0.1495	0.1745	0.1869	0.1931	0.1963
	Posterior	0.0001	0.1655	0.7218	0.0998	0.0117	0.0012
BIC		11.429	-2.906	-5.543	-1.448	2.913	7.420

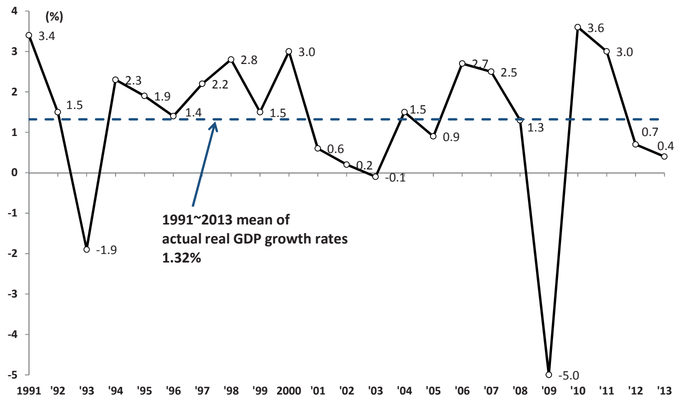
Notes: BIC=Bayesian Information Criteria

Table 7: Final model — combination of combinations(C^2)

	Prior: $P(C_j) = 1/6$	Prior: $P(C_1) < \cdots < P(C_6)$
dependent variable(y)	Germany's real GDP growth rate (yearly, %)	
regressors		
constant	0.572 (0.349)	0.576 (0.348)
RWI	2.400* (0.322)	2.413* (0.318)
IMF	-1.172* (0.248)	-1.168* (0.247)
EU	-0.610 (0.442)	-0.627 (0.438)
SVR	-0.033 (0.238)	-0.036 (0.249)
OECD	0.002 (0.057)	0.002 (0.061)
GD	-0.000 (0.030)	-0.000 (0.032)
RMSE	0.617	0.616
Effective Number of explanatory variables	2.926	2.961

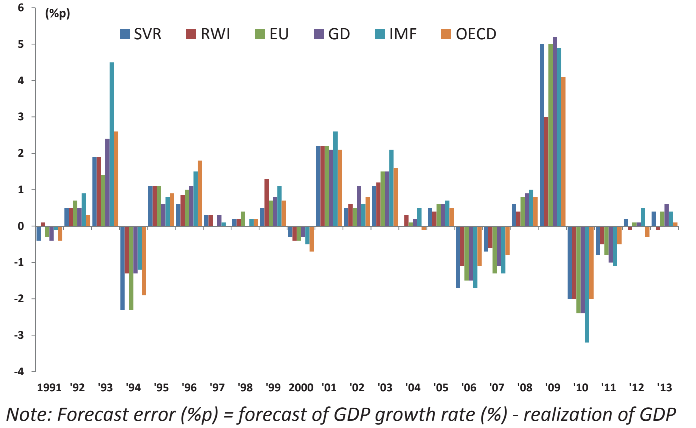
*Notes: standard errors in parentheses. * denotes p<0.05.*

Figure 1: Germany's real GDP growth rates, 1991~2013



Note: These rates are 'real time' data, rather than subject to methodological revisions such as transition to ESA1995.

Figure 2: Forecast errors in Germany's GDP growth rates, 1991~2013



Note: Forecast error (%p) = forecast of GDP growth rate (%) - realization of GDP growth rate (%)

Figure 3: Absolute values of forecast errors in %p, 2000~2013

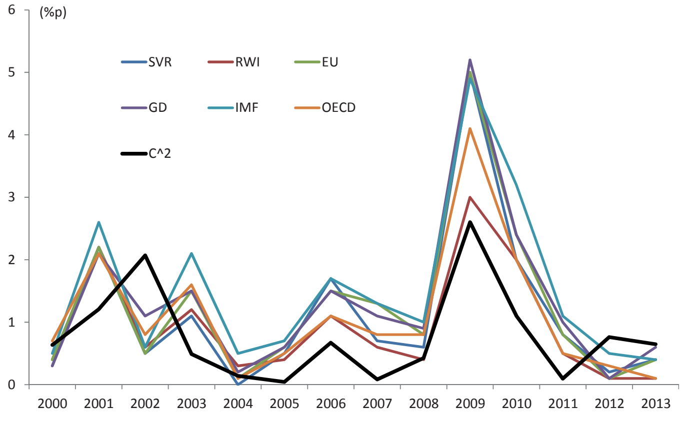
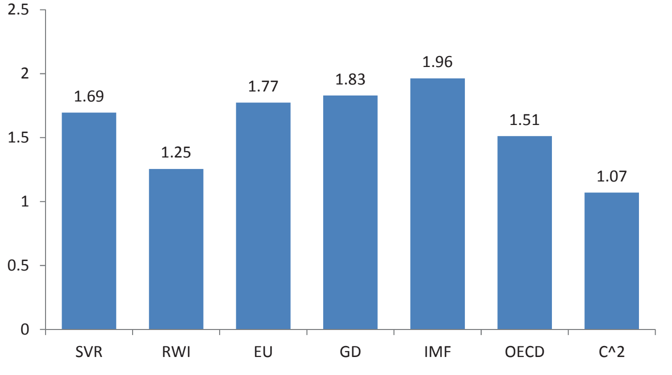


Figure 4: Root-mean-square prediction errors(RMSPE), 2000~2013



Note: As was in Table 1, RMSPE(root-mean-square prediction error) = $\sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - g_t)^2}$*, where* x_t *is a forecast and* g_t *is the realized value of year* t*'s real GDP growth rate.*