Crowdsourcing of Economic Forecast: Combination of Combinations of Individual Forecasts Using Bayesian Model Averaging

Tae-hwan Rhee and Keunkwan Ryu

Economic forecasts are essential in our daily lives. Accordingly, we ask the following questions: (1) Can we have an improved prediction when we additionally combine combinations of forecasts made by various institutions? (2) If we can, then what method of additional combination will be preferred? We non-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 forecasting South Korea's real GDP growth rate, this new forecast dominates any single forecast in terms of root-mean-square prediction errors. When compared with simple linear combinations of forecasts, our method works as a "hedge" against prediction risks, avoiding the worst combination while maintaining prediction errors similar to those of the best combinations.

Keywords: Combination of combinations, Combination of forecasts, Bayesian model averaging

JEL Classification: C53, E37

Tae-hwan Rhee, Corresponding Author, Associate Professor, Department of Economics, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Korea. (E-mail): thrhee@sejong.ac.kr, (Tel): +82-2-6935-2474; Keunkwan Ryu, Professor, Department of Economics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea. (E-mail): ryu@snu.ac.kr.

[**Seoul Journal of Economics** 2021, Vol. 34, No. 1]
DOI: 10.22904/sje.2021.34.1.006

I. Introduction

Economic forecast has never been easy. It is a task to predict the future values of economic variables of interest, such as GDP growth rate, consumer price index, and balance of international payments. It requires professional knowledge, skills, and experience. Given that economic variables are determined through nearly an infinite number of interactions among billions of humans, making a correct and precise forecast is nearly impossible. The popular assumption ceteris paribus will almost always turn out wrong.

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

Given the aforementioned importance, many research institutions periodically make and publish forecasts of main economic indicators. Among these institutions are international organizations, such as the International Monetary Funds (IMF), Organisation for Economic Co-operation and Development (OECD), and European Union (EU). South Korea has government-related or public institutions, such as the Bank of Korea (BOK) and Korea Development Institute (KDI); and private institutions, such as LG Economic Research Institute (LGERI). Global investment banks, such as Morgan Stanley and Citi Group, also issue periodic economic forecasts with regular or irregular intervals.

To make forecasts, these institutions introduce their respective specific models and assumptions regarding the future behavior of economic agents (*e.g.*, household, business, and government). Differences in the models and assumptions will lead to differences in their forecasts. For this reason, we typically have multiple forecasts for the same target variable, such as South Korea's GDP growth rate in 2020. This situation leads to a problem of choice: Which forecast is more reliable than others?

To enhance the reliability of forecasts, one can think of combining individual forecasts to construct a new forecast. Undoubtedly, a linear combination of forecasts, including a simple average, is well-established to often produce smaller prediction errors than a single forecast. However, one cannot ex ante know which forecasts are to be included in this combination to obtain the best performing model. Moreover, there is a good chance of reducing, rather than enhancing, predictive accuracy even if we blindly combine noisy pieces of information. We may substantially need a systematic method of evaluating and weighting different forms of linear combinations.

The central questions of this study are as follows: (1) Can we have a superior prediction when we combine multiple forecasts by taking an additional combination of several combinations of individual forecasts? (2) If we can, what method of combination will be preferred? We introduce a means of applying the method of Bayesian model averaging to the issue of forecast combination. Bayesian model averaging is well-known in the literature (*e.g.*, Zellner (1971), Leamer (1978), Liang and Ryu (2003)). The current research is not interested in deriving any new model averaging method but in applying the Bayesian model averaging to combine different forecasting models. Given that a linear combination of forecasts is already a model in itself, the Bayesian model averaging leads to a non-linear "combination of combinations." We combine the different linear combinations of existing forecasts to form a new forecast, and the weights are given by Bayesian posterior model probabilities.

One may ask why a double combination (*i.e.*, combination of combinations) is different from a single combination. The reason is that a single linear combination of forecasts is a linear function of only the component forecasts, whereas a double combination (*i.e.*, a non-linear combination of linear combinations) is no longer a linear function of the component forecasts as the combining weights given to different forecasts are functions of component forecasts and target values (detailed in Section II, *C*).

Our "combination of combinations" is different from a simple combination and also useful in forecasting. When we apply this method to the forecasts of South Korea's GDP growth rates, combining the forecasts made by four 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). When compared with simple linear combinations of forecasts, our method works as a "hedge" against prediction risks. Our final model easily outperforms the worst performing combination and shows prediction errors similar to those of the best performing combinations. However, the proposed model cannot ex post beat every simple linear combination every year in terms of prediction accuracy.

The remainder of this paper is organized as follows. Section II introduces the model and method of Bayesian model averaging. Section III deals with the data used in this research. Section IV shows the application of our methodology using South Korea's GDP forecasts made by four different institutions and summarizes the results. Lastly, Section V concludes this research with implications and directions for further studies.

II. Model

A. Outline of the model

Let y_{t+1} be the variable of interest, such as the real GDP growth rate for the upcoming year $t + 1$. There are k different institutions making forecasts of y_{t+1}. We denote institute j's forecast of y_{t+1} at time t by $x_{t+1,j}$, $j = 1,\cdots,k$. For now, we only deal with one-period ahead forecasts of y_{t+1}, although it can be easily extended. Let I_t be the information set available at time t. Note that for each t, y_t, $x_{t+1,1},\cdots,x_{t+1,k}$ and their previous values are in the information set I_t but not y_{t+1}.

We assume that we are at T and want to forecast y_{T+1}. Given k different forecasts, $x_{t+1,1},\cdots,x_{t+1,k}$, which are available, consider constructing a model to form a new forecast of y_{T+1}. In this case, a model corresponds to a so-called "combination of forecasts." The model is as follows:

$$y_{T+1} = \beta_0 + \beta_1 x_{T+1,1} + \beta_2 x_{T+1,2} + \beta_k x_{T+1,k} + e_{T+1}. \tag{1}$$

Using $y_1,y_2,\cdots,y_T$ and $x_1,x_{2,j},\cdots,x_{T+1,j}(j = 1,2,\cdots,k)$ in the information set I_t, we would like to forecast y_{T+1}.

Our combination of combinations proceeds in two steps. In the first step, we estimate several interim models of the form in (1), resulting in combinations of forecasts. In the second step, we combine these interim models using Bayesian model averaging method, resulting in a combination of combinations (denoted as C^2 hereafter).

Each interim model can contain a different number of forecasts from

1 to k. If we order the forecasts in advance, then we may consider k different interim models; the first contains only the best forecast, the second is a combination of two best forecasts, *etc.* Alternatively, we may have up to a maximum of $2^k - 1$ different models, similar to Sala-i-Martin *et al.* (2004). That is, each forecast may or may not be used in a linear combination of forecasts, and at least one forecast should be included in any given combination. We do not think this alternative method adds considerably to the simple one because it would eventually lead to the combination of numerous noisy forecasts. This conjecture turns out to be true in our analysis with the South Korean GDP data in the following sections.[1]

With pre-ordering, we consider k different interim models: $C_1, C_2, \cdots, C_k$.

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

where y is a $T \times 1$ vector of realized target values, l is a $T \times 1$ vector of ones, and x_j is a $T \times 1$ vector of institute j's forecasts. To save on notations, assume that $x_1 \succ x_2 \succ \cdots \succ x_k$ in terms of additional predictive contribution, where $A \succ B$ means A is preferred over B. Thereafter, the interim Model j (C_j) uses only $x_1, x_2, \cdots, x_j$, and it has $j+1$ parameters to estimate, including the constant term.[2] For generality, let us denote the number of parameters in Model j as k_j.

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

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

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

[1] See footnotes 14 and 15 for the detailed results using all the possible interim combinations.

[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 up to unity.

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

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

Several methods of ordering the existing forecasts can be used:

(a) root-mean-square error (RMSE) or mean absolute percentage error,

(b) sequential/stepwise R^2 criteria (as in Liang and Ryu (1996)), or

(c) subjective judgment.

When we apply the C^2 method to South Korea's growth rate forecasts in Section IV, we use the sequential R^2 criteria. The final model C^2 will be a non-linear combination of interim combinations $C_1, C_2, \cdots, C_k$, and the resulting forecast will be the weighted average of the interim forecasts, which themselves are the linear combinations of individual forecasts.

B. Bayesian posterior on β

To obtain $\hat{\beta}$, the Bayesian estimator of β in the final model

$$C^2 : y = X\beta + e,^3$$

we first formulate a Bayesian posterior density function of the β conditional on the observed y. Following the Bayesian model averaging method, this Bayesian posterior can be written as follows[4]:

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

[3] This is the vectoral form of (1).

[4] See Zellner (1971), Leamer (1978), Sala-i-Martin *et al.* (2004), Hansen (2007), or Fragoso *et al.* (2018).

which is the weighted average of model-specific posteriors, with each model's weight being equal to the posterior model probability $P(C_j \mid y)$.

Note that under non-informative (diffuse) priors and *i.i.d.* normal assumptions on $e_1^j, \cdots, e_T^j$, each model-specific posterior

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

is given by $N\left(\hat{\beta}^j, Var\left(\hat{\beta}^j\right)\right)$, where

$$\hat{\beta}^j = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_j \\ \cdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

is the restricted OLS estimator of β^j from Model j

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

and

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

with $Var\begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_j \end{pmatrix}$ being the conventional OLS variance from

$$y = X_j \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_j \end{pmatrix} + e^j.$$

C. Bayesian posterior model probabilities $P(C_j \mid y)$

Given prior odds ratio $P(C_i) / P(C_j)$, as the data information increases (*i.e.*, $X_l'X_l \to \infty$ for each $l \in \{1,2,\cdots,k\}$), we approximately have the following posterior odds ratio for model C_i vis-à-vis model C_j with $i \neq j$[5]:

$$\frac{P(C_i \mid y)}{P(C_j \mid y)} \approx \frac{P(C_i)\, T^{-k_i/2} SSE_i^{-T/2}}{P(C_j)\, T^{-k_j/2} SSE_j^{-T/2}}, \tag{2}$$

where

T = number of observations used for model estimation;
$P(C_j)$ = prior model probability;
k_j = number of parameters in Model j ($k_j = j + 1$); and
SSE_j = sum of the squared errors from the OLS estimation of Model j.

From (2), the posterior odds ratio between the two models C_i and C_j is evidently a product of the following three terms:

(i) prior odds ratio: $P(C_i) / P(C_j)$;
(ii) penalty for lack of "parsimoniousness": $T^{-k_i/2} / T^{-k_j/2}$; and
(iii) penalty for lack of "in-sample performance": $SSE_i^{-T/2} / SSE_j^{-T/2}$.

We use (2) to derive the posterior model probability up to a proportionality constant as follows:

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

Using the well-known Bayesian information criterion (BIC)

$$BIC_j = k_j \log T + T \log SSE_j, \tag{3}$$

we can rewrite the posterior model probability as follows:

[5] See Zellner (1971), Leamer (1978), or Hansen (2007).

$$P(C_j \mid y) \propto P(C_j)\exp\left[-\frac{1}{2}BIC_j\right].$$

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 $P(C_j)$. Let us consider a generalized prior generating scheme (just a form but not an absolute standard):

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

using a real number $\omega \in [0,1]$. In the following sections, we will try two values (*i.e.*, $\omega = 0$ and $\omega = 0.5$) and compare the results to each other. Note that "$\omega = 0$" corresponds to an equal prior across different models:

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

This uniform prior probability distribution is considered the standard specification. Note that it gives equal prior probability to each model and not to each forecast.[6] In the case of $\omega = 0.5$, the prior model probabilities are provided by the following equation:

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

which assigns higher prior probability to models containing larger number of forecasts.[7] This assignment scheme appears reasonable, considering that each forecast contains valuable information. However,

[6] Among the k different forecasts, x_1 is contained in all k models, but x_k is used only in the k_{th} model C_k. If we assign equal prior probability to every model, then we are in fact assigning the largest and lowest weights on x_1 and x_k, respectively.

[7] 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. Thus, we are assigning larger weight on the models using more forecasts, but not in strict proportion to the number of forecasts contained in each model.

we should also consider the possibility of adding noise rather than information when we blindly combine additional forecasts into a model. That is why we try $\omega = 0$ and $\omega = 0.5$ in the later sections. It turns out that the posterior model probability is not considerably sensitive to the choice of the prior model probability distribution.

Once we specify the model prior $P(C_j)$, we obtain the following equation:

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

D. Inferences on β based on $g(\beta \mid y)$

We have shown that (i) the Bayesian posterior density function $g(\beta \mid y)$ can be written as the expectation of the model-specific posteriors weighted by posterior model probabilities, and (ii) the model posteriors are determined by model priors and model selection criteria BIC_j.[8]

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

$$E\left(\beta \mid y\right) = \sum_{j=1}^{k} P\left(C_j \mid y\right) E\left(\beta \mid C_j, y\right) = \sum_{j=1}^{k} P\left(C_j \mid y\right) \hat{\beta}^j \tag{5}$$

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

$$\hat{\beta}^j = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_j \\ \cdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

[8] One may use alternative model selection criteria for BIC_j, such as the Akaike information criterion (AIC) or an increasing transformation of Mallows' C_p:
(i) $BIC_j = k_j \log T + T\log SSE_j$;
(ii) $AIC_j = 2k_j + T\log SSE_j$; and
(iii) $T\log C_p$, where $C_p = 2\sigma^2 k_j + SSE_j$ is the so-called Mallows' C_p.

is given by the OLS estimation of C_j and is equal to $E(\beta \mid C_j, y)$.

Using the variance decomposition formula[9],

$$Var\left(\beta \mid y\right) = E_{C_j} Var\left(\beta \mid C_j, y\right) + Var_{C_j}\left(E\left[\beta \mid C_j, y\right]\right) \tag{6}$$

$$= \sum_j P\left(C_j \mid y\right) Var\left(\beta \mid C_j, y\right) + \sum_j P(C_j \mid y)\left(\hat{\beta}^j - E\left(\beta \mid y\right)\right)\left(\hat{\beta}^j - E\left(\beta \mid y\right)\right)',$$

where the model-specific variance $Var(\beta \mid C_j, y)$ can be estimated as follows:

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

where the quantity in the northwestern block is the conventional OLS variance estimator from model C_j. We use (5) and (6) to make posterior inference on β. For example, we can compute a posterior 95% probability interval for β_j as $\sum_{i=1}^{k} P\left(C_i \mid y\right) \cdot \left[\hat{\beta}^i\right]_j \pm 2SE(\beta_j \mid y)$.[10]

III. Data

We apply the methodology introduced in the previous section to a real-world economic forecast. The target variable is South Korea's real GDP growth rate in percentage term (*e.g.*, 6.5 in 2000 and 2.0 in 2019). To build a correct series of this variable, note that there were multiple occasions of changes in South Korea's GDP accounting method. For example, BOK made the transition to chain-index pricing system in 2009, and they adopted a new set of guidelines to comply with the 2008 System of National Accounts (SNA) in 2014. We use the "real time" data of

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

[10] For the individual coefficient β_j on the j_{th} forecast

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

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

Table 1

Descriptive Statistics, 2000–2019 Forecasts and realization of South Korea's GDP growth rates in %

	BOK	KDI	IMF	OECD	Realized Values
Mean	4.10	4.20	4.14	4.13	3.79
Standard Deviation	1.09	1.25	0.99	1.11	1.84
RMSPE	1.41	1.35	1.86	1.61	

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

GDP growth rates, which were calculated under the accounting method considered current and official by the forecasting institutions when they published the forecasts.

We use forecasts by four institutions: BOK and KDI (domestic institutions) and IMF and OECD (international organizations). The timespan is from 2000 to 2019. Given that we use annual forecasts, we have 20 annual observations.

We also need to pay detailed attention to the timing of forecasts. The four institutions publish their growth forecasts at least twice a year. Thus, we should choose among multiple forecasts for each institution. For the forecast of GDP growth rate in year $t + 1$, we use the latest forecast available at the end of December in year t for each institution. For example, for South Korea's real GDP growth rate forecast for 2019, we take each institution's latest forecast available on December 31, 2018.

Table 1 summarizes the descriptive statistics of South Korea's real GDP growth rate forecasts from the four institutions, and the realized values. Moreover, Table 4 shows a few interesting characteristics. First, the mean of forecasts of each institution is higher than the mean of the actual growth rates. Since 2000, the four institutions have made more optimistic average forecasts than the actual status of the South Korean economy, with the gaps in the range of 0.3%p–0.4%p.

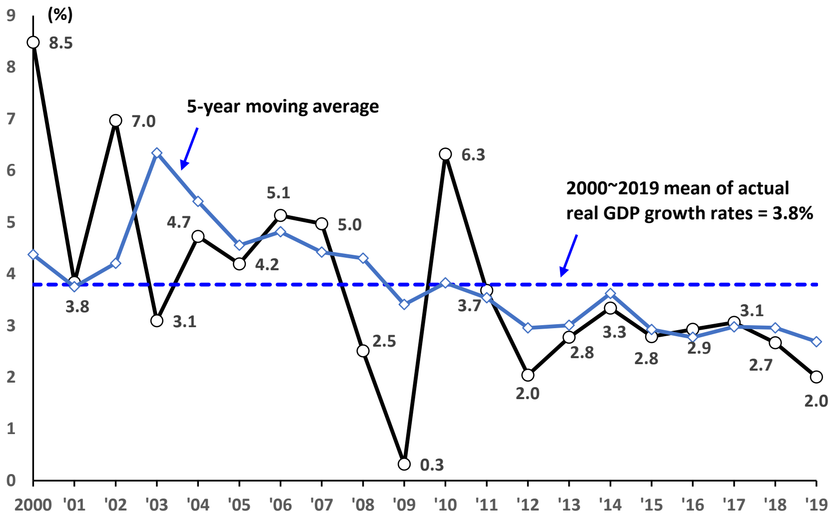
Second, the standard deviation of forecasts for each institution is significantly smaller than that of actual growth rates. The standard deviation of the realized values is 1.84, whereas that of the four forecasts are clustered around 1, with KDI's 1.25 being the largest. Thus, only approximately half to two-thirds of the movements of the

Table 2

Correlation coefficients among growth forecasts and realization

	KDI	IMF	OECD	Realized Values
BOK	0.916	0.844	0.955	0.672
KDI		0.747	0.906	0.715
IMF			0.910	0.284
OECD				0.523

Note: Bilateral correlation coefficients among the forecasts of South Korea's real GDP growth rates and the realized values.



Note: These rates are "real time" data, rather than subject to methodological revisions, such as transition to the chain-index system.

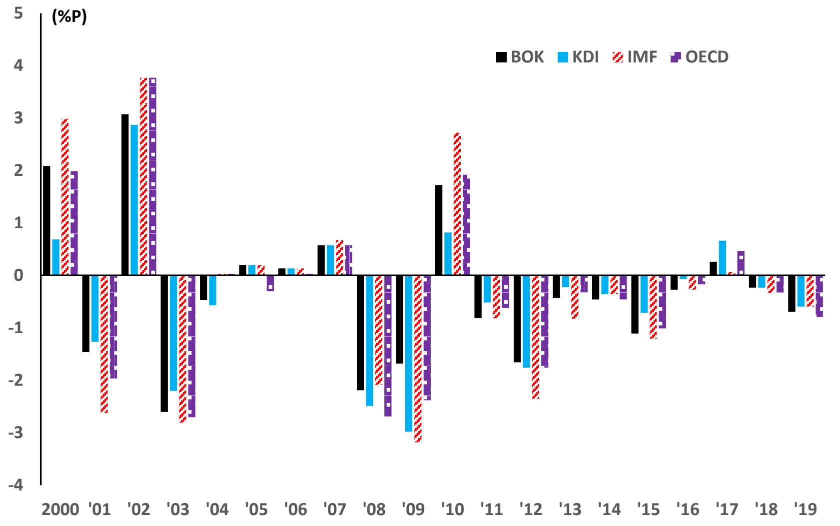
Figure 1

South Korea's real GDP growth rates, 2000–2019

actual growth rates are reflected in the forecast series. That is, the forecasts of the four institutions can be characterized as "conservative."

This conservatism is easily revealed when one compares Figures 1 and 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 South Korea's real GDP growth rates and the realized



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

Figure 2

In-sample forecast errors in South Korea's GDP growth rates, 2000–2019

values. Note that the correlation coefficients are positive among the four institutions; in general, the forecasts have moved in the same direction. The correlation between BOK and OECD is particularly high at 0.955, whereas that between KDI and IMF is relatively lower at 0.747. The four forecasts have been moving in similar directions during the 20-year period, but the specific direction and magnitude of movements vary annually.

A wider variation exists among the bilateral correlation coefficients between the forecasts and realization than among those between the forecasts. KDI's forecast shows a relatively high correlation (*i.e.*, 0.715) with the actual growth rate, whereas that of IMF shows considerably lower correlation (*i.e.*, 0.284).

In summary, considering RMSPE and the correlations between the forecasts and realization, KDI's forecasts have been the most precise, with the smallest RMSPE and highest correlation coefficient. IMF's forecast by itself can be evaluated as the least accurate, with the largest

RMSPE and lowest correlation.[11]

IV. Results

We use the methodology described in Section II as a tool and South Korea's GDP growth rate forecasts summarized in Section III as data to make a new "combination of combinations"(C^2) forecast based on the forecasts of the four institutions. Our goal is to show that we can form a new and more informative forecast if we make a suitable non-linear combination of combinations of forecasts and assign proper weight to each combination according to the method of Bayesian model averaging. This procedure comprises two steps, the first of which is to order the forecasts according to the in-sample fitting performance.

In the first round of the first step, we run four regressions in total. For each regression, we try each one of the four forecasts as the single explanatory variable, while the realized value of South Korea's GDP growth rate remains the dependent variable. The coefficient of determination (R^2) is highest when KDI's forecast is used. Equivalently, the in-sample root-mean-square error (RMSE) is lowest when using KDI's forecast. That is, KDI's forecast has the largest explanatory power in the case of a one-variable regression.[12] We denote these one-variable regression models as Models a, b, c, and d. The results of the first-round regressions are summarized in Table 3.

Given that KDI'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 two-variable regressions, three of them in total. For each regression KDI's forecast and one of the other three institutions' forecasts are used as explanatory variables.

Table 4 shows the largest gain in R^2 when IMF's forecast is added in

[11] We have to consider the differences in the timing of forecasts. IMF's forecasts are published in September, while those of KDI are published in November. Hence, KDI has a significant advantage in terms of accuracy. However, IMF's early forecast, or the difference between IMF and KDI's forecasts, contains beneficial information, as presented in the next section.

[12] Table 3 shows that the actual number of regressors is two, including the constant term, rather than one. We will continue to call this round as "one-variable regressions," emphasizing that "1" is the number of forecasts used in the regression equations.

TABLE 3
STEPWISE REGRESSION, ONE-VARIABLE CASES

	Model a	**Model b**	Model c	Model d
Dependent variable	Korea's real GDP growth rate (annually, %)			
Regressors				
Constant	−0.872 (1.255)	**−0.627** **(1.062)**	1.614 (1.783)	0.210 (1.427)
BOK	1.139* (0.296)			
KDI		**1.054*** **(0.243)**		
IMF			0.527 (0.419)	
OECD				0.868* (0.334)
R^2	0.451	**0.512**	0.081	0.273
RMSE	1.366	**1.288**	1.768	1.572

Notes: Standard errors in parentheses. * denotes $p < 0.05$.

this round.[13] 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 extremely low, with the coefficient of determination being a mere 0.081. However, working as an additional variable given KDI's forecast, IMF's forecast is actually the most informative in predicting the succeeding year's growth rate (largest marginal contribution to predictability).

Now we fix KDI and IMF's forecasts as explanatory variables, and a forecast from one of the two remaining forecasts is included as the third explanatory variable. In this round with two different three-variable regressions, one leading to the largest coefficient of determination is selected. At this point, we have ordered the four institutions according to the (additional) predictive powers of their forecasts (*i.e.*, KDI, IMF, BOK, and OECD).

As a result of the previous stepwise regressions, we now have four ($k = 4$) interim models, with each one being a combination of forecasts. The first model includes only KDI's forecast as an explanatory variable.

[13] Equivalently, Model bc shows the smallest RMSE.

TABLE 4
STEPWISE REGRESSION, TWO-VARIABLE CASES

	Model ba	**Model bc**	Model bd
Dependent variable	Korea's real GDP growth rate (annually, %)		
regressors			
Constant	−0.758 (1.218)	**1.106 (1.130)**	0.258 (1.089)
BOK	0.174 (0.715)		
KDI	0.916 (0.621)	**1.679* (0.316)**	1.994* (0.535)
IMF		**−1.053* (0.398)**	
OECD			−1.169 (0.602)
R^2	0.513	**0.654**	0.600
RMSE	1.286	**1.085**	1.166

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)
Dependent variable	South Korea's real GDP growth rate (annually, %)			
Regressors				
Constant	−0.627 (1.062)	1.106 (1.130)	0.918 (0.961)	0.531 (0.950)
KDI	1.054* (0.243)	1.679* (0.316)	0.692 (0.448)	1.039* (0.480)
IMF		−1.053* (0.398)	−1.749* (0.422)	−1.100 (0.574)
BOK			1.761* (0.639)	2.383* (0.726)
OECD				−1.526 (0.960)
R^2	0.512	0.654	0.765	0.799
Adjusted R^2	0.485	0.613	0.721	0.746
RMSE	1.288	1.085	0.893	0.826

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)
$P\left(C_j\right)=\frac{1}{4}$ ($\omega = 0$)	Prior	0.2500	0.2500	0.2500	0.2500
	Posterior	0.0242	0.0957	0.5657	0.3144
$P\left(C_j\right) \propto 1+\frac{1}{2}+\cdots+\left(\frac{1}{2}\right)^{j-1}$ ($\omega = 0.5$)	Prior	0.1633	0.2449	0.2857	0.3061
	Posterior	0.0139	0.0821	0.5666	0.3374
BIC		18.231	15.483	11.929	13.104

Notes: BIC = Bayesian information criteria

Hereinafter, we call KDI Institute 1, and denote this model as C_1 which reads "combination one." IMF becomes Institute 2, and the interim model is denoted as C_2 when it includes forecasts by Institutes 1 and 2, and so on. The regression results of $C_1, C_2, \cdots, C_4$ are summarized in Table 5. We also report the adjusted R^2 for each interim model, thereby confirming that the four institute's forecasts enhance the model fit.

For the second step of the procedure, the first round aims to evaluate the Bayesian posterior model probability for each of the four interim models. For this, we need to specify the prior probabilities, which can be done in many different ways as mentioned in Section II. We consider two cases: equal weighting scheme (ω = 0) and non-equal weighting scheme (ω = 0.5) which assigns a higher prior probability to a model with more explanatory variables. Once a model prior is specified, the posterior model probability $P(C_j \mid y)$ can be derived using (3) and (4). Table 6 compares the prior and posterior model probabilities for each interim model.

Regardless of the model prior, the highest posterior model probability is assigned to Model 3 (C_3). Except for Models 3 and 4, no other model is assigned a higher posterior probability than its prior. Model 4's posterior probability is only slightly higher than its prior. When making forecasts on South Korea's annual GDP growth rates from 2000 to 2019, Model 3, which linearly combines three forecasts made by Institutes 1, 2, and 3, is preferred among the four interim models. That Model 3 is the

Table 7
Final model; Combination of combinations (C^2)

	Prior: $P\left(C_j\right)=\frac{1}{4}$	Prior: $P\left(C_j\right)\propto 1+\frac{1}{2}+\cdots+\left(\frac{1}{2}\right)^{j-1}$
Dependent variable	South Korea's real GDP growth rate (annually, %)	
Regressors		
Constant	0.777 (1.022)	0.782 (1.008)
KDI	0.904 (0.535)	0.895 (0.530)
IMF	−1.436* (0.612)	−1.449* (0.600)
BOK	1.745 (0.943)	1.802* (0.914)
OECD	−0.480 (0.890)	−0.515 (0.912)
RMSE	0.862	0.858
Effective number of explanatory variables	3.170	3.228

Notes: Standard errors in parentheses. * Denotes $p < 0.05$.

most preferred is not sensitive to the way prior model probabilities are assigned.

However, selecting Model 3 is not our final destination. The second round of the current second step aims to combine the four models using the weights given by the posterior model probabilities. This procedure of combining the combinations (C^2) will give our final forecast model. Inserting the model-by-model regression results in Table 5 and posterior probabilities in Table 6 into Formulas (4), (5), and (6) gives us the final set of results (*i.e.*, expected values and standard errors of the regression coefficients 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. RMSEs of these final models are 0.862 and 0.858, which are relatively lower than the 1.288 of Model 1 (C_1). Model 1's explanatory variable is Institute 1's forecast, which has the largest explanatory power among the one-variable regressions using only one institute's forecast as an explanatory variable. Therefore, our final model (C^2) shows better performance than any single institution's forecast with regard to in-sample fitting.[14]

Evidently, we have to focus on the number of explanatory variables when we try this type of interpretation. In regression analyses, a higher number of explanatory variables mechanically leads to a smaller RMSE within the sample period. Thus, our final model, which utilizes multiple forecasts, can be naturally expected to have a smaller RMSE than any other model using only one forecast. To properly evaluate the in-sample performance of our final model vis-à-vis a single forecast or a simple combination, we should consider the "effective" number of explanatory variables (ENEV) in our final model, which we define as follows:

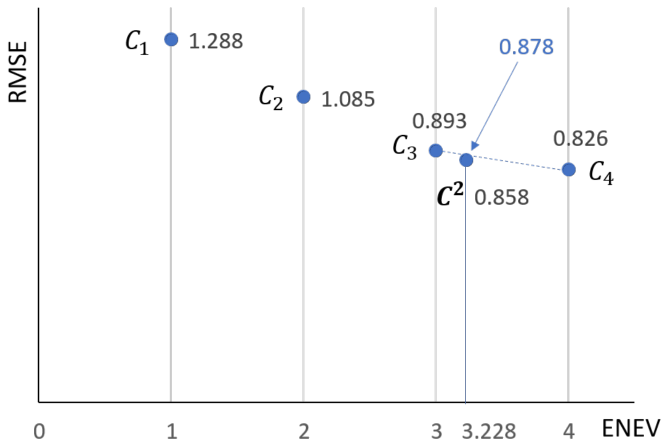
$$\text{ENEV} \equiv \sum_{j=1}^{k} j \cdot P(C_j \mid y)$$

Except for constants, Model 1 (C_1) has one effective explanatory variable *i.e.*, ENEV = 1, and Model 2's ENEV is two. In the case of combination of combinations, ENEV is defined as the weighted average of the ENEVs of the component combination models with weights given by the posterior model probabilities. The reason is that a combination-of-combinations model is a weighted average of the component interim models. In the case of increasing priors of $\omega = 0.5$ ($P(C_1) < P(C_2) < \cdots P(C_4)$), the effective number of explanatory variables can be calculated as follows:

$$0.0139 \times 1 + 0.0821 \times 2 + \cdots + 0.3374 \times 4 = 3.228$$

We evaluate again the in-sample fitting performance of the final model in terms of RMSE, specifically by considering the notion of ENEV. In the case of increasing priors, ENEV is 3.228 and RMSE is 0.858. This RMSE is quite smaller than 1.288, the RMSE of Model 1 (C_1) with ENEV = 1. The RMSE of C^2 is also below 0.893, the RMSE of Model 3 (C_3) with

[14] We also try our method of double combination using the 15 (= $2^4 - 1$) interim models. That is, we try Bayesian model averaging with every possible linear combination of forecasts from the four institutions, as in Sala-i-Martin *et al.* (2004). The resulting final model C^2_{all} shows poorer performance than the ones combining only the four interim models. In the cases of uniform and increasing prior model probability distributions, RMSEs of C^2_{all} are 0.870 and 0.865, respectively.



Note: ENEV = Effective number of explanatory variables.

Figure 3

Linear interpolation of C^2's in-sample RMSE between C_3 and C_4's RMSE's

ENEV=3, and relatively above 0.826, the RMSE of Model 4 (C_4) with ENEV = 4. This result 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 3.228 is between 3 and 4, if we go from 0.893 to 0.826 by 22.8%, then we will be at 0.878, which is slightly above 0.858 from C^2. Figure 3 graphically shows this interpolation on the ENEV-RMSE plane.

By nature of forecasting, out-of-sample performance is a more important criterion than in-sample performance. Our data period is not considerably long, thereby preventing us from meaningfully comparing out-of-sample performance across different forecasts. Nevertheless, we try to evaluate the out-of-sample prediction performance of our final model using whatever data available to us.

Figure 4 shows the absolute values of forecast errors from 2010 to 2019. Here, the forecasts of our final model (C^2) are calculated recursively. For example, for 2010's forecast, we estimate the coefficients of C^2 using the forecasts by the four institutions and realization of the GDP growth rates in 2000-2009, then we insert the four institutions' 2010 forecasts into this C^2. The 2011–2019 forecasts of C^2 are calculated in the same, recursive way. The "increasing" prior model probabilities

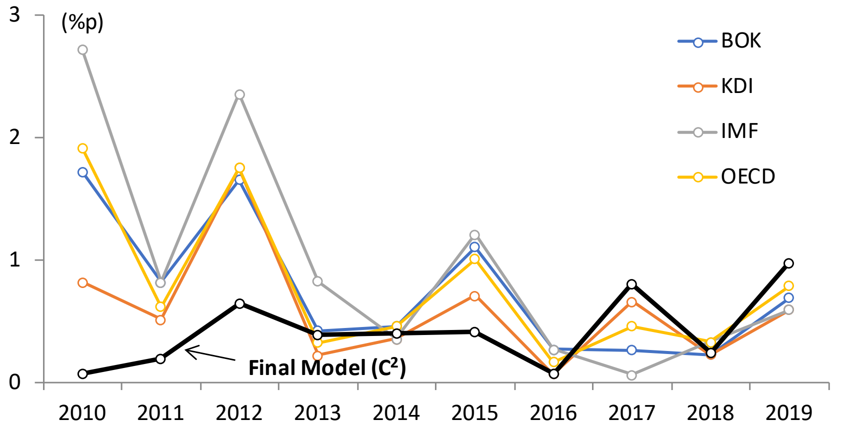
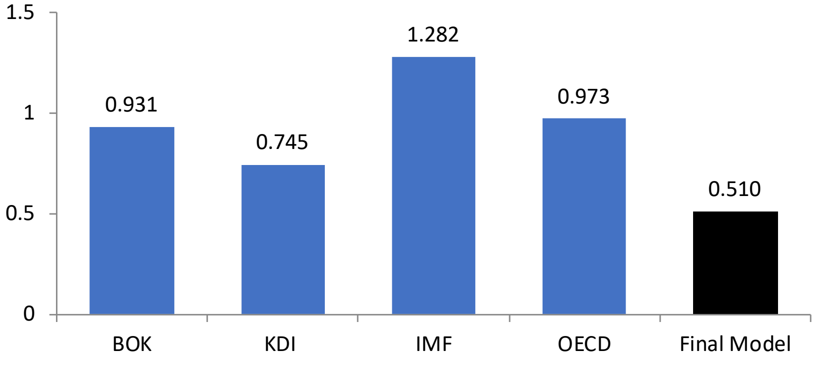


Figure 4

Absolute values of out-of-sample forecast errors in %p, 2010–2019



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

Figure 5

Root-mean-square prediction errors, 2010–2019

($\omega = 0.5$) are used for this exercise.

Note that in terms of absolute forecast errors, C^2 beats any individual institution's forecast in 4 out of 10 years (*i.e.*, 2010, 2011, 2012, and 2015) from 2010 to 2019. Among the four years, 2010–2012 is the period immediately after the global financial crisis, and 2015 is when South

Table 8
Out-of-sample prediction performance measures, 2010–2019

Model	Model 1 (C_1)	Model 2 (C_2)	Model 3 (C_3)	Model 4 (C_4)	Final model (C^2)
Average Rank	3.30	3.20	3.50	2.50	**2.50**
MAPE	0.62	0.49	0.52	0.44	**0.42**

Notes: MAPE = Mean absolute prediction error. Each model's forecast is calculated in a recursive way, annually, from 2010 to 2019. For the average ranks, the model with the smallest absolute prediction error is ranked 1st each year.

Korea was hit by the Middle East respiratory syndrome (MERS) epidemic. Although these shocks have led to unusually inaccurate forecasts, the out-of-sample performance of C^2 can still be considered impressive. Figure 5 compares RMSPEs of the forecasts shown in Figure 4 for the entire out-of-sample period (*i.e.*, 2010–2019). RMSPE of C^2 is 0.510, which easily beats any single institution's forecast in the 10-year period.

We compare the out-of-sample prediction performance of C^2 with that of the interim models, $C_1,\cdots,C_4$. Here the results are not unambiguous. That is, the performance of C^2 is sufficiently good but cannot beat every single interim model. To provide more detailed information, we present two additional measures, namely, average rank and mean absolute prediction error (MAPE), apart from RMSPE. These measures are summarized in Table 8.[15]

In terms of the average ranks, C_4 ties with C^2. In 2010–2019, when we rank the absolute prediction errors among the five models ($C_1,\cdots,C_4$ and C^2) each year, the average rank of both C_4 and C^2 is 2.5.[16] For MAPEs, in which not only the ranks but also the sizes of prediction errors matter, C^2 beats every single interim model. Lastly, in terms of RMSPEs, in which we punish larger errors more severely, C^2 is ranked second among the five models, beating $C_1,\cdots,C_3$ but not C_4. Figures 6 and 7 show the absolute values of the forecast errors and RMSPEs, respectively, of the five models.

[15] When we combine the 15 possible interim models with increasing prior model probabilities, RMSPE and MAPE of the final model C^2_{all} are 0.589 and 0.51, respectively. These values are larger than those of our final model C^2 with only four interim models.

[16] For each year, the model with the smallest absolute prediction error is ranked first, while that with the largest error is ranked fifth.

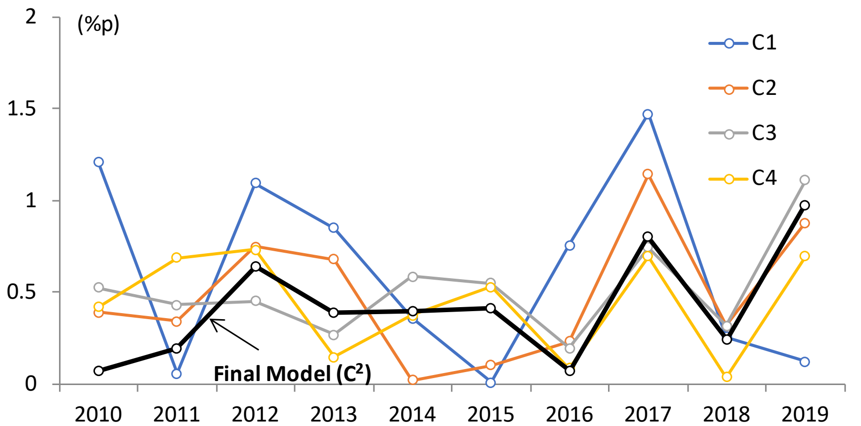
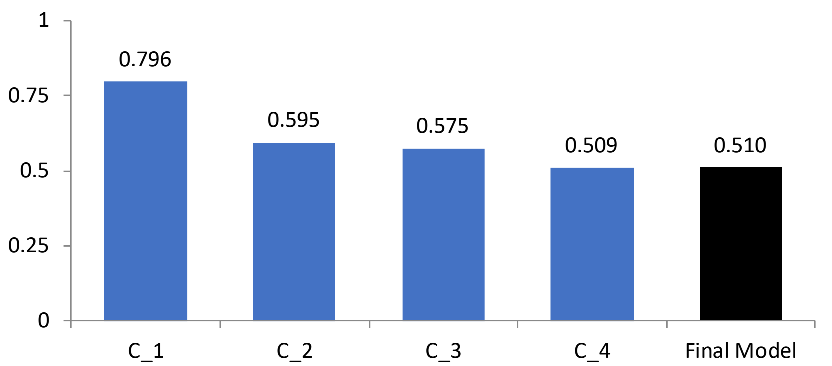


Figure 6
Absolute values of out-of-sample forecast errors in %p, 2010–2019



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

Figure 7
RMSPE, 2010–2019

In terms of all three measures (i.e., average rank, MAPE, and RMSPE), overall C^2 appears to be the best performing model. Even under the RMSPE criterion, the performance of C^2 is considerably close to that of C_4 and easily beats the other combination models. These results suggest that C^2 can be considered a "hedge" against prediction risk, while enjoying superior forecasting performance overall. Note that we cannot ex ante know which interim model will turn out to be the best in out-of-sample forecasting. That is, we do not know which forecasts will improve the prediction accuracy when they are included in a simple combination of forecasts. In this study, the best performing interim combination C_4 beats the final model C^2 with a small gain in accuracy. However, we do not know whether this will be true in other cases, with some other target variables and forecasting institutions in a different period. Our method can construct a forecast that easily outperforms inferior combinations, and provides a superior, substantially robust forecast.

V. Conclusion

The following question is the starting point of our research: Can we make a new, more precise forecast when we combine multiple existing combinations of forecasts?[17] We have first shown that the method of Bayesian model averaging could be applied as a useful weighting scheme. We have constructed multiple linear models, and evaluated the posterior model probabilities of these interim models according to Bayesian theory. Our final model, which we call "combination of combinations", or C^2, is the combination of the interim models using the posterior model probabilities as the weights. Note that our combination of combinations is a non-linear combination of combinations, and thus does not reduce to a linear combination of forecasts. Hence, we denote our method as C^2 and not as C.

Against this theoretical background, we have applied our method to the forecasts of South Korea's GDP growth rates made by four different institutions. The final model we have derived beats any single forecast in terms of RMSPE for 2010–2019. When compared with simple linear combinations, the final model works as a "hedge" against prediction risk, outperforming the inferior combinations and showing prediction errors similar to those of the best combinations. Although the data length is not long, we have a favorable signal that our method could

[17] Note that as emphasized in Section I, linear combinations of forecasts are linear functions of only the components forecasts. By contrast, non-linear combinations of linear combinations are no longer linear functions of the component forecasts.

actually be used to improve the precision of economic forecasts by combining multiple existing forecasts and/or multiple forecasting methods.

Lastly, note that our method has a wide range of applicability. The C^2 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, such as current account balances, international oil prices, and stock market indices. We are optimistic to see numerous applications and further studies.

(Received 9 September 2020; Revised 17 January 2021; Accepted 25 January 2021)

References

Fragoso, T., W. Bertoli and F. Louzada. "Bayesian Model Averaging: A Systematic Review and Conceptual Classification." *International Statistical Review* 86 (No.1 2018): 1-28.

Granger, C. and R. Ramanathan. "Improved Methods of Combining Forecasts." *Journal of Forecasting* 3 (No.2 1984): 197-204.

Hansen, B. "Least Squares Model Averaging." *Econometrica* 75 (No.4 2007): 1175-1189.

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

Leamer, E. "Let's Take the Con Out of Econometrics." *The American Economic Review* 73 (No.1 1983): 31-43.

Liang, K. and K. Ryu. "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, 1996.

Liang, K. and K. Ryu. "Relationship of Forecast Encompassing to Composite Forecasts with Simulations and an Application." *Seoul Journal of Economics* 16 (No.3 2003): 363-386.

Sala-i-Martin, X., G. Doppelhofer and R. I. Miller. "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." *The American Economic Review* 94 (No.4 2004): 813-835.

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

Zellner, A. "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: 233-243. Amsterdam: North-Holland, 1986.