Optimal Forecasting Groups

P. J. Lamberson, Scott E. Page,

MANAGEMENT SCIENCE
Vol. 58, No. 4, April 2012, pp. 805–810
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.1110.1441

Optimal Forecasting Groups

P. J. Lamberson
Kellogg School of Management and Northwestern University Institute on Complex Systems, Northwestern University, Evanston, Illinois 60208, p-lamberson@kellogg.northwestern.edu

Scott E. Page
Center for the Study of Complex Systems, Departments of Economics and Political Science, University of Michigan, Ann Arbor, Michigan 48106, spage@umich.edu

This paper characterizes the optimal composition of a group for making a combined forecast. In the model, individual forecasters have types defined according to a statistical criterion we call *type coherence.* Members of the same type have identical expected accuracy, and forecasters within a type have higher covariance than forecasters of different types. We derive the optimal group composition as a function of predictive accuracy, between- and within-type covariance, and group size. Group size plays a critical role in determining the optimal group: in small groups the most accurate type should be in the majority, whereas in large groups the type with the least within-type covariance should dominate.

Key words: combining forecasts; optimal groups; information aggregation
History: Received July 15, 2010; accepted May 25, 2011, by Peter Wakker, decision analysis. Published online in *Articles in Advance* October 28, 2011.

1. Introduction

Combining multiple forecasts is a well-accepted approach for reducing forecasting error (Clemen 1989, Armstrong 2001). An investment firm may combine financial models to predict the future price of a commodity, or individual members of a marketing team may aggregate their diverse forecasts of a new product's expected sales. Given a collection of forecasts, a body of research from finance, forecasting, operations research, and machine learning describes how to best create an aggregate prediction based on information about the relative accuracy and similarity (covariance) of the individual predictions.[1] The problem has been sufficiently well studied that some of the formal results have been translated into simple rules of thumb, such as "use at least five forecasts when possible," and "use equal weights unless you have strong evidence to support unequal weighting of forecasts" (Armstrong 2001, pp. 420, 422).

The question left unexplored, which this paper takes up, is which forecasts should be combined. In other words, who should be part of the forecasting group? Not surprisingly, more accurate constituent forecasts lead to a more accurate combined forecast. Less intuitively, research demonstrates that combinations of forecasts of different types (i.e., that rely on different approaches) also improve performance (Ashton 1986, Batchelor and Dua 1995, Armstrong 2005, Cuzán et al. 2005, Lobo and Nair 2007). However, the most accurate forecasts may well be based on similar models and assumptions, creating a tradeoff between accuracy and diversity (Hong and Page 2004). To characterize the optimal group composition, this paper splits forecasters into categories, called types, based on a statistical criterion that we call *type coherence*. The criterion implies that the designated categories have statistical relevance. Specifically, forecasts within a type will be more highly correlated than forecasts made by different types, on average. For example, in the case of forecasting student improvement, forecasters might be divided into two types: one that focuses on teacher attributes and a second that emphasizes attributes of the students.

The group size and the distribution across types determines the combined forecast's expected accuracy. The main result shows how the optimal group composition depends on four features: the *accuracies* of the types, the *within-type covariances*, the *across-type covariances*, and the *group size*. Of particular note, this paper proves that as group size increases the accuracy of a type matters less than its within-type covariance. If we interpret low within-type covariance as high within-type model diversity, this result implies that larger groups should be comprised of types that are internally more diverse even if those types are not as accurate. In small groups, the opposite holds. The most accurate type should be in the majority.

[1] See Bates and Granger (1969), Reid (1969), Newbold and Granger (1974), Winkler (1981), Granger and Ramanathan (1984), Clemen and Winkler (1986), Clemen (1989), Kang (1986), Schmittlein et al. (1990), and Armstrong (2001).

2. Optimal Weighting of Models

We consider a forecasting problem in which a collection of individuals must predict a cardinal value, V. The problem could be predicting the value of a firm, the sales of a new product, or the vote share of a political candidate. This section summarizes well-known results on optimal aggregation for a given set of forecasts that serve as a baseline of comparison for our results on group composition (Markowitz 1952, Bates and Granger 1969, Winkler 1981, Tresp and Taniguchi 1995, Ueda and Nakano 1996). Here, optimal means minimizing the mean squared forecasting error, which is equivalent to minimizing the error variance in the case of unbiased forecasts.

Assume there are M forecasters. Following convention, the predictions of these M forecasters are modeled as signals represented by real valued random variables. Denote the prediction/forecast/signal of forecaster i by s_i. Let $\varepsilon_{s_i} = s_i - V$ denote the error of signal s_i, and let $\text{var}(\varepsilon_{s_i})$ denote the variance of the error. The higher the variance, the less accurate the prediction. Finally, let Σ denote the covariance matrix of the errors in the signals. The covariance between two signals is a measure of their diversity. In practice Σ must be estimated, typically from past predictions. As is standard in much of the combining literature, Σ is assumed to be both known and stable over time.

Given a vector of signals $s = (s_1, \ldots, s_M)$, the group prediction is a real valued random variable $G(s)$, where $G\colon \mathbb{R}^M \to \mathbb{R}$. We restrict our attention to weighted averages:

$$G(s) = w_1 s_1 + w_2 s_2 + \cdots + w_M s_M, \tag{1}$$

with

$$\sum_{i=1}^{M} w_i = 1. \tag{2}$$

The most common aggregation method in practice seems to be simple averaging. This aggregation rule is optimal only when all of the forecasts are independent and equally accurate. Nonetheless, it has been advocated as a good operational rule of thumb for many situations (Makridakis and Winkler 1983, Clemen and Winkler 1986, Clemen 1989, Schmittlein et al. 1990, Armstrong 2001). The expected squared error of a simple average can be written as the familiar *bias–variance–covariance* decomposition:

Theorem 1. *The expected squared error of $G(s)$ when G gives the simple average of the signals $s_1, \ldots, s_M$ is*

$$\mathrm{E}[(G(s) - V)^2] = \overline{\text{bias}}(s)^2 + \frac{1}{M}\overline{\text{var}}(s) + \left(1 - \frac{1}{M}\right)\overline{\text{cov}}(s), \tag{3}$$

where

$$\overline{\text{bias}}(s) = \frac{1}{M}\sum_{i=1}^{M}(\mathrm{E}[\varepsilon_{s_i}]) \tag{4}$$

denotes the average bias of the signals,

$$\overline{\text{var}}(s) = \frac{1}{M}\sum_{i=1}^{M}\text{var}(\varepsilon_{s_i}) \tag{5}$$

denotes the average variance of the signal errors, and

$$\overline{\text{cov}}(s) = \frac{1}{M(M-1)}\sum_{i=1}^{M}\sum_{j\neq i}\text{cov}(\varepsilon_{s_i}, \varepsilon_{s_j}) \tag{6}$$

denotes the average covariance of the signal errors.

A straightforward calculation allows one to similarly decompose the expected squared error of any weighted average of the individual signals. Given this decomposition, the optimal weights to place on the various signals can be derived as shown in Theorem 2.

Theorem 2. *The optimal weighted average of a collection of unbiased signals $s = (s_1, \ldots, s_M)$ with error covariance matrix Σ is given by*

$$G(s) = (u'\Sigma^{-1}s)/(u'\Sigma^{-1}u), \tag{7}$$

where $u = (1, \ldots, 1)'$ is the $M \times 1$ unit vector. The expected squared error of this weighted average is

$$\mathrm{E}[(G(s) - V)^2] = (u'\Sigma^{-1}u)^{-1}. \tag{8}$$

Example 1 (Two Signals). Suppose that two unbiased signals, a and b, have error covariance matrix

$$\Sigma = \begin{bmatrix} \text{var}(\varepsilon_a) & \text{cov}(\varepsilon_a, \varepsilon_b) \\ \text{cov}(\varepsilon_a, \varepsilon_b) & \text{var}(\varepsilon_b) \end{bmatrix}. \tag{9}$$

Then the optimal weights on a and b given by Theorem 2 are

$$\frac{\text{var}(\varepsilon_b) - \text{cov}(\varepsilon_a, \varepsilon_b)}{\text{var}(\varepsilon_a) + \text{var}(\varepsilon_b) - 2\,\text{cov}(\varepsilon_a, \varepsilon_b)} \tag{10}$$

and

$$\frac{\text{var}(\varepsilon_a) - \text{cov}(\varepsilon_a, \varepsilon_b)}{\text{var}(\varepsilon_a) + \text{var}(\varepsilon_b) - 2\,\text{cov}(\varepsilon_a, \varepsilon_b)}, \tag{11}$$

respectively. Note that if the errors in the two signals are independent, then the weight attached to each signal is inversely proportional to the relative error variance of that signal. If the errors of signal b have twice the variance of signal a, then signal b receives one half the weight of signal a.

3. Optimal Group Composition

In the remainder of this paper we assume that each forecaster has a type. Intuitively, types can be thought of as different approaches to the forecasting problem at hand. Formally, types are defined by a statistical criterion, *type coherence*, described below. For clarity of exposition, we assume there are only two types of forecasters, type a and type b, but the results extend to groups of arbitrarily many types.

3.1. Type Coherence

Assume that all forecasts made by type a forecasters have expected squared error $\text{var}(\varepsilon_a)$, and forecasts made by type b forecasters have expected squared error $\text{var}(\varepsilon_b)$ (assume the forecasters are unbiased). Let $\text{cov}(\varepsilon_a)$ denote the covariance in the errors of the forecasts made by two different type a forecasters, and assume that this is the same for any pair of type a forecasters. Define $\text{cov}(\varepsilon_b)$ similarly. Finally, let $\text{cov}(\varepsilon_a, \varepsilon_b)$ denote the covariance in the errors of any two forecasters, one of which is of type a and the other of which is of type b.

Definition 1. Two types, a and b, satisfy *type coherence* if $TC(a, b) > 0$, where

$$TC(a, b) = \text{cov}(\varepsilon_a) + \text{cov}(\varepsilon_b) - 2\,\text{cov}(\varepsilon_a, \varepsilon_b). \tag{12}$$

In words, the types satisfy type coherence if the between-type covariance is less than the average of the within-type covariances. In what follows, we assume type coherence unless otherwise mentioned. Most of the results require only the type coherence assumption, but at times we require a stronger notion of types:

Definition 2. Two types a and b satisfy *strong type coherence* if

$$\begin{gathered} \text{cov}(\varepsilon_a) > \text{cov}(\varepsilon_a, \varepsilon_b) \quad \text{and} \\ \text{cov}(\varepsilon_b) > \text{cov}(\varepsilon_a, \varepsilon_b). \end{gathered} \tag{13}$$

Type coherence can be verified using past prediction data. The extent of error covariance within a type depends on the breadth of that class of models along with the quality of information available. The covariance across types depends on how much they rely on common variables and the differences in the methods used to form their estimates.

3.2. Group Forecasts

We assume that the group forecast is formed by taking a simple average of the individual forecasts. We make this restrictive assumption for several reasons. First, the focus of this paper is the optimal group composition rather than how to combine a given set of forecasts. Second, this assumption keeps the resulting formulas as transparent as possible so the qualitative insights can be more easily extracted from the equations. Third, the simple average is the most commonly employed method in practice, so it seems practical to condition our recommendations on this premise rather than on the less realistic assumption that the forecasts will be combined using the optimal weights. Fourth, experts recommend the simple average in many real world settings (Clemen and Winkler 1986, Clemen 1989, Kang 1986, Schmittlein et al. 1990, Armstrong 2001).

From Theorem 1, the expected squared error of a simple average of a group with A type a forecasters and B type b forecasters is given by the following expression, where $M = A + B$:

$$\frac{A\text{var}(\varepsilon_a) + B\text{var}(\varepsilon_b) + 2\binom{A}{2}\text{cov}(\varepsilon_a) + 2\binom{B}{2}\text{cov}(\varepsilon_b) + 2AB\text{cov}(\varepsilon_a, \varepsilon_b)}{M^2}. \tag{14}$$

Given the size of the group, M, we can solve for the number of type a forecasters that minimizes this expected error.

Theorem 3. *In a group of size M, the optimal fraction of type a forecasters is approximated by*

$$\frac{[\text{var}(\varepsilon_b) - \text{var}(\varepsilon_a)] - [\text{cov}(\varepsilon_b) - \text{cov}(\varepsilon_a)]}{2M \cdot TC(a, b)} + \frac{\text{cov}(\varepsilon_b) - \text{cov}(\varepsilon_a, \varepsilon_b)}{TC(a, b)}, \tag{15}$$

where the approximation error is less than $1/M$.[2]

Theorem 3 makes it clear that decreasing $\text{var}(\varepsilon_a)$ or increasing $\text{var}(\varepsilon_b)$ implies that the optimal group should contain more type a forecasters, regardless of the other parameters. The effects of changes in group size, $\text{cov}(\varepsilon_a)$, $\text{cov}(\varepsilon_b)$, or $\text{cov}(\varepsilon_a, \varepsilon_b)$, are conditional on the other parameter values. For example, when the a forecasts become more similar to one another, that is, $\text{cov}(\varepsilon_a)$ increases, the optimal group contains more type a forecasters if and only if

$$\frac{\text{var}(\varepsilon_a) - \text{var}(\varepsilon_b)}{2(M-1)} < \text{cov}(\varepsilon_b) - \text{cov}(\varepsilon_a, \varepsilon_b). \tag{16}$$

As group size increases, accuracy plays an increasingly small role.

3.3. Large Group vs. Small Group Composition

As the group size goes to infinity, we have the following theorem:

Theorem 4. *As the group size M approaches infinity, the optimal fraction of type a forecasters approaches*

$$\frac{\text{cov}(\varepsilon_b) - \text{cov}(\varepsilon_a, \varepsilon_b)}{TC(a, b)}. \tag{17}$$

[2] The proof follows from treating (14) as a function of A for all $A \in \mathbb{R}$. The first-order condition identifies M times (15) as a critical point, and positive type coherence guarantees that the critical point is a minimum. This minimum may not be an integer, but the minimum of (14) restricted to the integers must be one of the two nearest integers that bracket this value. Thus, the error of the approximation for the optimal number of A forecasters is less than one, and the error for the optimal fraction of A forecasters is less than $1/M$. If forecasters within a type are identical, so that $\text{cov}(\varepsilon_a) = \text{var}(\varepsilon_a)$ and $\text{cov}(\varepsilon_a) = \text{var}(\varepsilon_a)$, the first term vanishes, and we recover the optimal weight from Equation (10).

The limiting fraction depends on both the within- and between-type covariances but does not depend on the accuracies of either the type a or the type b forecasters. For large groups diversity matters, but accuracy does not.

Corollary 1. *In a sufficiently small group, the lowest variance type should be in the majority. In a sufficiently large group, the forecaster type with the lowest within-type covariance should be in the majority.*

A natural question is how large or how small of a group is needed in Corollary 1. One can easily solve for the group size required to guarantee that the optimal group contains at least or at most 50% of a given forecaster type. The formula is especially clear when the two types are assumed to be independent.

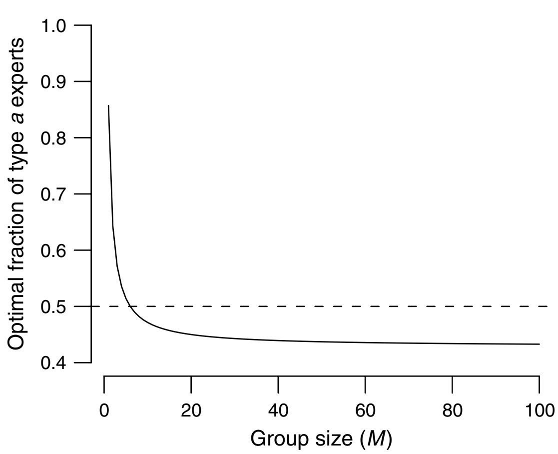
Theorem 5. *Suppose that the forecasts of type a and type b forecasters are independent of one another, i.e.,* $\mathrm{cov}(\varepsilon_a, \varepsilon_b) = 0$, *and that* $\mathrm{cov}(\varepsilon_b) < \mathrm{cov}(\varepsilon_a)$. *Then a group of size M should contain more than 50% type b forecasters if*

$$\frac{\mathrm{cov}(\varepsilon_a) - \mathrm{cov}(\varepsilon_b)}{\mathrm{var}(\varepsilon_b) - \mathrm{var}(\varepsilon_a)} < \frac{1}{M+1}. \tag{18}$$

Thus, the threshold ratio of the difference between within-type covariances and the type variances increases linearly in group size; if the group size becomes twice as large, the ratio of differences in covariances to differences in accuracies needs only be half as large.

Suppose, for example, that $\mathrm{var}(\varepsilon_a) = 5$, $\mathrm{var}(\varepsilon_b) = 10$, $\mathrm{cov}(\varepsilon_a) = 2$, $\mathrm{cov}(\varepsilon_b) = 1$, and $\mathrm{cov}(\varepsilon_a, \varepsilon_b) = -2$. Then an optimal small group contains more type a forecasters, but an optimal large group contains more type b forecasters. Figure 1 illustrates the phenomenon. In an optimal group of size six or less, at least 50% of the forecasters are of type a, but in an optimal group of more than six forecasters, at least half of them are of the less accurate type b.

Figure 1 **Small Groups Prize Accuracy, Large Groups Prize Diversity**



Note. The optimal fraction of type a forecasters in a group of size M from Theorem 3 with $\mathrm{var}(\varepsilon_a) = 5$, $\mathrm{var}(\varepsilon_b) = 10$, $\mathrm{cov}(\varepsilon_a) = 2$, $\mathrm{cov}(\varepsilon_b) = 1$, and $\mathrm{cov}(\varepsilon_a, \varepsilon_b) = -2$ is shown.

3.4. Group Composition vs. Weights

One might expect that the optimal fraction of each type of forecaster in a large group would equal the optimal weights for a weighted average of only two forecasters. We show this not to be the case. In fact, in some cases it is optimal to place higher weight on the type a forecaster, but when taking the simple average of a large population it would be optimal to have more type b forecasters.

First, suppose there is one type a forecaster and one type b forecaster. Recall from Equation (10) that the optimal weight on a type a forecaster equals

$$\frac{\mathrm{var}(\varepsilon_b) - \mathrm{cov}(\varepsilon_a, \varepsilon_b)}{\mathrm{var}(\varepsilon_a) + \mathrm{var}(\varepsilon_b) - 2\,\mathrm{cov}(\varepsilon_a, \varepsilon_b)}. \tag{19}$$

Next, suppose that we have a large group. From Theorem (17), the optimal fraction of type a forecasters approaches

$$\frac{\mathrm{cov}(\varepsilon_b) - \mathrm{cov}(\varepsilon_a, \varepsilon_b)}{TC(a, b)} = \frac{\mathrm{cov}(\varepsilon_b) - \mathrm{cov}(\varepsilon_a, \varepsilon_b)}{\mathrm{cov}(\varepsilon_a) + \mathrm{cov}(\varepsilon_b) - 2\,\mathrm{cov}(\varepsilon_a, \varepsilon_b)}. \tag{20}$$

Comparing Equations (19) and (20), we can see that the role of variance in the optimal weights of Equation (19) is played by within-type covariance in the optimal fraction of Equation (20).

Consider again the case depicted in Figure 1 with $\mathrm{var}(\varepsilon_a) = 5$, $\mathrm{var}(\varepsilon_b) = 10$, $\mathrm{cov}(\varepsilon_a) = 2$, $\mathrm{cov}(\varepsilon_b) = 1$, and $\mathrm{cov}(\varepsilon_a, \varepsilon_b) = -2$. The optimal weight on the type a forecast in a group consisting of one type a forecaster and one type b forecaster is $12/19 \approx 0.632$, but the optimal fraction of type a forecasters as the size of the population approaches infinity converges to $3/7 \approx 0.429$. Again, the reason for this switch in relative weighting is that when weighting models, only accuracy and between-type covariance matter, but when choosing a composition, covariance within types plays an important role.

4. Applications

This section considers three questions relevant to management decisions about who should participate in forecasting and about the use of collective subjective predictions.

4.1. Question 1: Are Two Types Better Than One?

When is it worthwhile to take team positions away from the most accurate type of forecaster in favor of less accurate forecasters of a different type? To answer this question, we examine when a group of only type a forecasters outperforms a group consisting of type a and type b forecasters.

Theorem 6. *The optimal group of size M contains both type a and type b forecasters if and only if*

$$TC(a,b) - M[\text{cov}(\varepsilon_b) - \text{cov}(\varepsilon_a,\varepsilon_b)] \le F(a,b) \le M[\text{cov}(\varepsilon_a) - \text{cov}(\varepsilon_a,\varepsilon_b)] - TC(a,b), \quad (21)$$

where

$$F(a,b) := \frac{[\text{var}(\varepsilon_b) - \text{var}(\varepsilon_a)] - [\text{cov}(\varepsilon_b) - \text{cov}(\varepsilon_a)]}{2}. \quad (22)$$

The theorem follows from requiring the optimal fraction of type a forecasters given in Theorem 3 to lie between $1/M$ and $(M-1)/M$. The left inequality in Equation (21) guarantees that the optimal group has at least one type a forecaster, and the right-hand inequality implies that the optimal group contains at least one type b forecaster.

The condition (21) becomes more intuitive under the assumption of strong type coherence. If strong type coherence holds, then for sufficiently large groups (21) is always satisfied. Thus, if types are sufficiently meaningful, then large groups should always be inclusive of both types. Though statistically straightforward, this finding runs counter to how many people think. It states that, provided types differ (the strong type coherence assumption), the optimal forecasting group should include members of the less accurate type.

4.2. Question 2: Include a Group Forecast?

The second question asks when should a group forecast be used to supplement a single model. To answer this question, set $B = 1$ in Equation (14). Then the expected squared error for an aggregate forecast composed of one type b forecast and A type a forecasts is

$$\frac{A\,\text{var}(\varepsilon_a) + \text{var}(\varepsilon_b) + A(A-1)\text{cov}(\varepsilon_a) + 2A\,\text{cov}(\varepsilon_a,\varepsilon_b)}{(A+1)^2}. \quad (23)$$

As A increases, this error approaches $\text{cov}(\varepsilon_a)$, so if $\text{cov}(\varepsilon_a) < \text{var}(\varepsilon_b)$, then including the forecast from a sufficiently large type a group reduces the expected squared error from the single type b forecast. On the other hand, if $\text{cov}(\varepsilon_a) \ge \text{var}(\varepsilon_b)$, then for large enough type a groups, including the group forecast along with the single type b prediction will increase the expected error. Put less formally, including a large group forecast will be beneficial if and only if the group is sufficiently internally diverse relative to the accuracy of the single model. The optimal decision rule depends on both the relationship between $\text{cov}(\varepsilon_a)$ and $\text{var}(\varepsilon_b)$ and whether or not the error (23) has a local minimum or maximum (as a function of A). If

$$\frac{\text{var}(\varepsilon_b) - \text{var}(\varepsilon_a)}{2} \le \text{cov}(\varepsilon_a) - \text{cov}(\varepsilon_a,\varepsilon_b), \quad (24)$$

then (23) has a local minimum. Assuming (24) holds, if $\text{var}(\varepsilon_b) \ge \text{cov}(\varepsilon_a)$, then including a type a group forecast always reduces the expected error, regardless of the group size. If (24) holds and $\text{var}(\varepsilon_b) < \text{cov}(\varepsilon_a)$, then including the type a group forecast reduces the expected error if and only if the number of type a forecasts, A, is less than

$$\tilde{A} = \frac{2\,\text{var}(\varepsilon_b) - \text{var}(\varepsilon_a) + \text{cov}(\varepsilon_a) - \text{cov}(\varepsilon_a,\varepsilon_b)}{\text{cov}(\varepsilon_a) - \text{var}(\varepsilon_b)}. \quad (25)$$

In this case, the optimal size for the type a group is

$$A^* = \frac{2\,\text{var}(\varepsilon_b) - \text{var}(\varepsilon_a) + \text{cov}(\varepsilon_a) - \text{cov}(\varepsilon_a,\varepsilon_b)}{3\,\text{cov}(\varepsilon_a) - 2\,\text{cov}(\varepsilon_a,\varepsilon_b) - \text{var}(\varepsilon_a)}. \quad (26)$$

If the condition (24) does not hold, then (23) has a local maximum. In this case, if $\text{var}(\varepsilon_b) \le \text{cov}(\varepsilon_a)$, then it is never advantageous to include the type a group forecast. If $\text{var}(\varepsilon_b) > \text{cov}(\varepsilon_a)$, then including the type a group forecast is beneficial if and only if the group has more than $\tilde{A}$ members.

4.3. Question 3: Add a Single Model of Another Type?

Finally, we consider when a single person with a different type of model should be added to a collective that consists of a single type. This captures situations in which an organization that uses a collection of econometric models to make a forecast were to also include the outcome from a prediction market. It also captures situations in which an outsider—someone with a different type of model—is added to the mix of forecasters. To answer this question, we compare the accuracy of a collection of A type a models with that of a group consisting of A type a models and a single type b model.

The expected squared error for the combined group is again given by (23), but now this error is compared to the expected error of the type a group. It follows that the combined group outperforms the type a only group if and only if

$$\frac{A\,\text{var}(\varepsilon_a) + \text{var}(\varepsilon_b) + A(A-1)\,\text{cov}(\varepsilon_a) + 2A\,\text{cov}(\varepsilon_a,\varepsilon_b)}{(A+1)^2} \le \frac{\text{var}(\varepsilon_a) + (A-1)\,\text{cov}(\varepsilon_a)}{A}. \quad (27)$$

This reduces to

$$2A^2[\text{cov}(\varepsilon_a,\varepsilon_b) - \text{cov}(\varepsilon_a)] + A[\text{var}(\varepsilon_b) - 2\,\text{var}(\varepsilon_a) + \text{cov}(\varepsilon_a)] - \text{var}(\varepsilon_a) \le 0. \quad (28)$$

For large groups, the quadratic term dominates the inequality, so the only consideration is the comparison of $\text{cov}(\varepsilon_a, \varepsilon_b)$ and $\text{cov}(\varepsilon_a)$. If the new forecast resembles the members of the group less than the group members resemble each other, then it should be included.

5. Conclusion

The main result of this paper characterizes optimal group composition as a function of type accuracy, within-type covariance, across-type covariance, and group size. For small groups, accuracy dominates, and the group should consist primarily of the most accurate types. For large groups, an opposite result holds. Within-type covariance, a proxy for model diversity, matters most. The application of these theoretical results in practice requires knowledge of the statistical properties of the forecasters. The natural question to investigate empirically is what causes a class of models to be more or less positively covariant.

The results provide insights into which forecast combinations are likely to be most effective. For example, combining two forecasts of approximately equal accuracy will almost always be beneficial. This insight held true in the Netflix Prize competition (Netflix 2009). At the end of the competition, the top two teams, BelKor's Pragmatic Chaos and The Ensemble, each improved on the Cinematch benchmark by 10.1%. After the competition, these two teams' models (themselves both combinations of many previous rival teams' models) were combined using simple averaging and produced a 10.5% improvement. (Keep in mind that in this competition improvements were measured in hundredths or thousandths of a percent.)

Given advances in information technology, the results have practical significance that they might not have had even a decade ago. Many companies now use prediction markets to aggregate the forecasts of large numbers of people. The results suggest that when combining many opinions, covariance within a type of forecaster matters more than the accuracy of those predictions: the less positively covariant a type, the more of those forecasters the group should include. Thus, prediction markets that can draw from large populations may do well to draw from types with high internal diversity.

Acknowledgments

This research was supported by a National Science Foundation IGERT Grant and the Army Research Office [Grant W911NF1010379].

References

Armstrong, J. S. 2001. Combining forecasts. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic, Norwell, MA, 417–439.

Armstrong, J. Scott. 2005. The forecasting canon: Nine generalizations to improve forecast accuracy. *Foresight* **1**(1) 29–34.

Ashton, R. H. 1986. Combining the judgments of experts: How many and which ones? *Organ. Behav. Human Decision Processes* **38**(3) 405–414.

Batchelor, R., P. Dua. 1995. Forecaster diversity and the benefits of combining forecasts. *Management Sci.* **41**(1) 68–75.

Bates, J. M., C. W. J. Granger. 1969. The combination of forecasts. *Oper. Res. Quart.* **20**(4) 451–468.

Clemen, R. T. 1989. Combining forecasts: A review and annotated bibliography. *Internat. J. Forecasting* **5**(4) 559–583.

Clemen, R. T., R. L. Winkler. 1986. Combining economic forecasts. *J. Bus. Econom. Statist.* **4**(1) 39–46.

Cuzán, A. G., J. S. Armstrong, R. Jones. 2005. Combining methods to forecast the 2004 presidential election: The Pollyvote. Presentation, Southern Political Science Association, Statesboro, GA.

Granger, C. W. J., R. Ramanathan. 1984. Improved methods of combining forecasts. *J. Forecasting* **3**(2) 197–204.

Hong, L., S. E. Page. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc. Natl. Acad. Sci. USA* **101**(46) 16385–16389.

Kang, H. 1986. Unstable weights in the combination of forecasts. *Management Sci.* **32**(6) 683–695.

Lobo, G. J., R. D. Nair. 2007. Combining judgmental and statistical forecasts: An application to earnings forecasts. *Decision Sci.* **21**(1) 446–460.

Markowitz, H. 1952. Portfolio selection. *J. Finance* **7**(1) 77–91.

Makridakis, S., R. L. Winkler. 1983. Averages of forecasts: Some empirical results. *Management Sci.* **29**(9) 987–996.

Netflix. 2009. The Netflix Prize. Accessed July 1, 2010, http://www.netflixprize.com.

Newbold, P., C. W. J. Granger. 1974. Experience with forecasting univariate time series and the combination of forecasts. *J. Roy. Statist. Soc. Ser. A* **137**(2) 131–165.

Reid, D. J. 1969. A comparative study of time series prediction techniques on economic data. Ph.D. thesis, University of Nottingham, Nottingham, UK.

Schmittlein, D. C., J. Kim, D. G. Morrison. 1990. Combining forecasts: Operational adjustments to theoretically optimal rules. *Management Sci.* **36**(9) 1044–1056.

Tresp, V., M. Taniguchi. 1995. Combining estimators using nonconstant weighting functions. G. Tesauro, D. S. Touretzky, T. K. Leen, eds. *Advances in Neural Information Processing Systems 7*. MIT Press, Cambridge, MA, 419–426.

Ueda, N., R. Nakano. 1996. Generalization error of ensemble estimators. *Proc. IEEE Internat. Conf. Neural Networks*, IEEE, New York, 90–95.

Winkler, R. L. 1981. Probability distributions from dependent information sources. *Management Sci.* **27**(4) 479–488.