A Randomization Rule for Selecting Forecasts

Dean P. Foster; Rakesh V. Vohra

Operations Research, Volume 41, Issue 4 (Jul. - Aug., 1993), 704-709.

Stable URL:
http://links.jstor.org/sici?sici=0030-364X%28199307%2F08%2941%3A4%3C704%3AARRFSF%3E2.0.CO%3B2-6

A RANDOMIZATION RULE FOR SELECTING FORECASTS

DEAN P. FOSTER
University of Pennsylvania, Philadelphia, Pennsylvania

RAKESH V. VOHRA
Ohio State University, Columbus, Ohio

(Received August 1988; revisions received August 1989, January 1991; accepted October 1992)

We propose a randomized strategy for selecting/combining forecasts that is better than the forecasts used to produce it in a sense made precise in this paper. Unlike traditional methods this approach requires that no assumptions be made about the distribution of the event being forecasted or the error distribution and stationarity of the constituent forecasts. The method is simple and easy to implement.

This paper is concerned with the problem of choosing from among a set of forecasts. It arises when forecasts from different methods, models, or sources are in conflict. One is then faced with the problem of choosing the best or more accurate forecast. We use the word forecast in a broad sense. We consider any statement about the future that is falsifiable in the sense of Popper (1968) to be a forecast. In other words, it should be possible to verify to the satisfaction of any observer that the forecast is either correct or in error. Furthermore, this error should be measurable.

A number of writers (Makridakis and Winkler 1983, Clemen and Winkler 1986, and Schnarrs 1984) have argued that selecting the best forecast, a priori, is usually difficult and perhaps impossible. For example, one approach to selecting a forecast is to identify the forecast that maximizes the decision maker's expected utility. This approach requires that the probability distribution of the event being forecasted be specified. Furthermore, even if this were known, the expected utility calculation may be quite burdensome. As a result, attention has shifted to methods for combining forecasts. Ideally, this means eviscerating the models behind each forecast, identifying their best features, and then combining them to form a new model and so a new forecast. (We point out that this is difficult and often impossible to do.) The argument is that a forecast that is produced by combining different forecasts will possess the best features of the individual forecasts. In this way, the combined forecast should outperform the individual forecasts. For an extensive survey of methods for combining forecasts the reader is referred to Clemen (1989). The most popular methods for combining forecasts involve taking a weighted average of the individual forecasts one has to choose from. The methods differ in how the weights are selected. Quite often the weights are selected to minimize some measure of forecast error. These minimizing approaches require that assumptions be made about the distribution of errors of the constituent forecasts and their stationarity. Also, like the expected utility approach described above, they can be computationally burdensome. For a succinct and critical review of these methods the reader should consult Gupta and Wilson (1987).

This paper proposes a new method for combining forecasts that is considerably more parsimonious in the assumptions it makes as well as the effort required to execute it. We call it the Mixing Method (MM for short). More importantly, MM is provably better, in a sense to be made precise later, than the individual forecasts used to produce it. We know of no other result of this kind in the literature on forecast combinations.

In the next section we make precise the notion of better alluded to above as well as introduce some of the notation to be used throughout the paper. Section 2 describes MM and discusses its properties, and Section 3 contains some modifications of MM. We conclude with an Appendix that contains proofs of all the results stated in the body of this paper.

1. A DEFINITION OF BETTER

Forecasters or forecasting methods will be denoted by capital letters. We assume that all forecasters have agreed upon a choice of error measure. By an error measure we mean a nonnegative real-valued function

Subject classification: Forecasting: combining forecasts.
Area of review: SIMULATION.

0030-364X/93/4104-0704 $01.25

of what was forecast as well as what transpired, which is nonincreasing with increasing accuracy of the forecast. The error measure of a particular (forecast, event) pair will simply be called the error of the forecast. The error of a forecaster in period i will be denoted by a lower case letter indexed by i. Thus, the error of forecast A in period i is a_i for example. The n period average error of a forecaster will be denoted by $T(A, n)$. So, for example

$$T(A, n) = \frac{1}{n} \sum_{i=1}^{n} a_i.$$

We list below two examples of error measures.

Example 1. Suppose that A is forecasting the demand for widgets. Then one choice of error measure might be:

a_i = |{the amount that A forecasted would be demanded in period i} − {the amount actually demanded in period i}|.

Alternatively, we might consider the square of the difference above.

Example 2. Suppose that A is forecasting the optimal objective function value of a minimization problem and L is a lower bound to the optimal solution value. Then one choice of a_i is:

a_i = {A's forecast of optimal objective function value} − L.

Suppose that A and B are two forecasters and C is a forecast that is "produced" from A and B. We will say that C is *better* than A and B if $T(C, n) \leq \min\{T(A, n), T(B, n)\} + \epsilon_n$, where $\epsilon_n \geq 0$ and $\epsilon_n \to 0$ in probability as $n \to \infty$. Informally, the n period average error of C is with high probability smaller than the n period average error of both A and B as $n \to \infty$. This definition of better encompasses two ideas. First, in comparing the accuracy of forecasts it is reasonable to examine their average error in order to assess their *typical* behavior. It therefore seems appropriate that C would be preferred to A or B if $T(C, n) \leq \min\{T(A, n), T(B, n)\}$. However, because of the nondeterministic nature of the problem it would be unreasonable to require that C be deemed better than A and B only if

$$T(C, n) \leq \min\{T(A, n), T(B, n)\}$$

for all n. After all,

$$T(C, n) > \min\{T(A, n), T(B, n)\}$$

for just one value of n is surely no reason to reject C out of hand. Thus, the milder requirement we impose in our definition of better.

In the next section we show how, given any two forecasts, it is possible to construct a third forecast that is better than both in the sense defined above.

2. THE MIXING METHOD

Let A and B be two forecasters. Denote by A_i and B_i their forecasts for period i, respectively. The Mixing Method constructs a forecaster C who operates as:

$$C_{n+1}$$
$$= A_{n+1} \text{ with probability } \min\left(\max\left[0, \frac{nD(n) + n^s}{2n^s}\right], 1\right)$$
$$= B_{n+1} \text{ otherwise,}$$

where $D(n) = T(B, n) - T(A, n)$ and $0.5 \leq s < 1$. We will say more about s later.

Note that the probability that MM picks A_{n+1} over B_{n+1} depends only on the previous history of A and B's errors. Furthermore, this probability changes as we move from one period to the next. We are now in a position to state the main results of the paper.

Theorem 1. *If A and B are two forecasters with a_i, b_i bounded for all i and C is defined by MM, then C is better than A and B.*

The number s affects the rate at which $\epsilon_n \to 0$ in Theorem 1. It can be seen from the proof of Theorem 1 in the Appendix that the optimal choice of s is 0.5. Notice that the only assumption being made is that the errors of A and B are bounded from above. This is a mild restriction because the quantities of most things that are forecasted are finite. For example, the amount of rain to fall on any day in any part of the world will never be less than zero inches and is unlikely to exceed 1,000 inches. If it did, it would be an act of God and unlikely to be *forecastable*! Apart from the requirement that errors are bounded, no assumption is made about the nature of what is being forecasted or how A and B come by their forecasts. In this sense, MM is a big departure from other approaches to combining forecasts based on minimizing a measure of forecast error or maximizing expected utility. These approaches require that assumptions be made about the distribution of errors of the constituent forecasts and their stationarity or the distribution of

the event being forecasted. As these assumptions rarely seem to hold, the use of such optimizing methods is limited.

MM is unusual in the sense that it randomizes between forecasts rather than the conventional approach of taking a weighted average of the individual forecasts. This randomization is a natural consequence of the scarcity of our assumptions. To see this, imagine that nature selects a probability distribution Φ for the event being forecasted. Our goal is to construct from A and B a forecast C that is better than either forecast. Recall now that we make no assumption about Φ. Hence, we have to construct a C better than A and B for every possible choice of Φ, including, of course, the most perverse and pathological. In effect, in designing C, one must imagine that one is in competition with nature.

It is natural to ask whether there is a version of MM that involves taking a weighted average of A and B for which the conclusion of Theorem 1 still holds. There is, provided one restricts oneself to error functions that are convex functions of |forecast − observed|. For specificity, suppose that we have two forecasters A and B forecasting the demand for widgets. Let W_i be the actual demand for widgets in period i. Denote by S the forecast obtained by taking a convex combination of the forecasts of A and B respectively, i.e., $S_i = \lambda_i A_i + (1 - \lambda_i)B_i$ for some $0 \leq \lambda_i \leq 1$. Let R be the forecast obtained by randomizing between A and B, i.e.,

$$R_i = A_i \text{ with probability } \lambda_i$$

$$= B_i \text{ with probability } 1 - \lambda_i.$$

Suppose that forecast error is measured according to $f(|\text{forecast} - \text{observed}|)$, where f is nonnegative, nondecreasing and convex (for example, $|\text{forecast} - \text{observed}|^2$) i.e., $a_i = f(|A_i - W_i|)$. Then,

$$s_i = f(|S_i - W_i|) = f(|\lambda_i A_i + (1 - \lambda_i)B_i - W_i|)$$

$$= f(|\lambda_i(A_i - W_i) + (1 - \lambda_i)(B_i - W_i)|)$$

$$\leq \lambda_i a_i + (1 - \lambda_i)b_i,$$

which is the average error from using R. Hence, in this situation the average error from randomizing is larger than the error from using a weighted average. Suppose that the λ_i's are chosen in accordance with MM. This would make R better than A and B and would immediately imply that S is also better than A and B. One can imagine an error function that is neither a convex or concave function of |forecast − observed|. Such error functions can arise from economic considerations. In such cases, it is not possible to make an argument like the one above to produce an averaging version of MM.

An instance of when randomizing is to be preferred to averaging is when the forecasts to be combined are correlated. As a concrete instance suppose that in the widget example mentioned earlier A and B have the property that A_i and B_i both exceed W_i (actually it is enough for this to happen most of the time). Suppose also that our decision maker has an error function that is concave and nondecreasing in |forecast − observed|. This amounts to saying the decision maker is a risk taker. Then, by an argument similar to the earlier widget example, we can prove that the average error of R is smaller than the error of S.

It is possible to extend MM so that it combines three or more forecasts to produce a better forecast without decreasing the rate at which $\epsilon_n \to 0$. However, the proof is long and tedious. We describe instead a quick scheme for combining forecasts that clearly produces a forecast that is better than the constituent forecasts. To combine three or more forecasts by MM we proceed iteratively. For example, suppose there are three forecasters, A, B, and E. First combine A and B using MM to get C. Then combine C and E using MM to get C'. It is easy to see that by virtue of Theorem 1, C' is better than A, B, and E. Note that this scheme is not associative. The lack of associativity means that it is possible to produce more than one forecast better than the constituent forecasts. This would be a problem if one sought the 'best' of the better forecasts. Because our objective is only to produce a better forecast this problem can be ignored.

Most of the work required to construct C from A and B is in the effort to update and record the errors of A and B. This can become quite burdensome if the cost of obtaining a forecast from A and B is high. In the next section, we show how to modify MM to avoid having to obtain a forecast from A and B in every period.

3. THE MODIFIED MIXING METHOD

Recall that the greatest investment of effort in determining C is in computing $D(n)$. We want to modify this to avoid having to get a forecast from A and B in every period from the first to the current one. We do this by using a statistical estimate of $D(n)$, which we call $\hat{D}(n)$, that is cheaper to compute.

Let r be a number between $2(1 - s)$ and 1 and $\{\alpha_i\}_{i=1}^n$ be a sequence of independent random variables uniformly distributed in [0, 1]. Let $\{X_i\}_{i=1}^n$ be a sequence of binomial random variables defined as: $X_i = 1$ if $\alpha_i \leq n^{r-1} = 0$ otherwise.

The variable X_i tells us whether to obtain a forecast from A and B in period i or not. If $X_i = 1$, then we get a forecast, otherwise we do not. We define $\hat{D}(n)$ to be

$$\frac{1}{n}\sum_{i=1}^{n}\frac{X_i(a_i - b_i)}{n^{r-1}}.$$

It should be clear that $\hat{D}(n)$ is cheaper than $D(n)$ to compute. The next result tells us that $\hat{D}(n)$ is both an unbiased and accurate estimator of $D(n)$. We assume throughout that a_i and b_i are bounded. Define Σ to be the σ-field generated by all infinite sequences $\{a_i\}$ and $\{b_i\}$.

Theorem 2. $E(\hat{D}(n)|\Sigma) = D(n)$ *and* $\mathrm{Var}(\hat{D}(n)|\Sigma) = 0(n^{-r})$.

The next theorem says that if we use $\hat{D}(n)$ in place of $D(n)$ in our construction of C the conclusion of Theorem 1 still holds. More precisely, if

$$\hat{C}_{n+1}$$
$$= A_{n+1} \text{ with probability } \min\left(\max\left[0, \frac{n\hat{D}(n) + n^s}{2n^s}\right], 1\right)$$
$$= B_{n+1} \text{ otherwise.}$$

Then:

Theorem 3. *If A and B are any two forecasts, then $\hat{C}$ is better than A and B.*

APPENDIX

To prove Theorem 1 we first need a purely technical lemma. Before stating this lemma we introduce some simplifying notation. Let $(x, y)^+ = \max(x, y)$ and $(x, y)^- = \min(x, y)$.

Lemma. *Let $F(t)$ be a real-valued function defined on the interval $[0, n]$, where n is a positive integer. Suppose that there is a function $f(s)$ such that*

i. $|f(s)| \leq 1 \ \forall s \in [0, n]$,
ii. $f(s) = f(\lfloor s \rfloor)$,
iii. $F(t) = \int_0^t f(s)\,ds$, $F(n) \leq 0$.

Then

$$I_n = \int_0^n \left[\left(0, \frac{F(t) + t^\alpha}{2t^\alpha}\right)^+, 1\right]^- f(t)\,dt = 0(n^\alpha),$$

where $0 < \alpha < 1$ is fixed.

Proof. Let $Z(t) = (F(t) + t^\alpha)/2t^\alpha$. Observe that $Z(n) < 1$. Hence

$$I_n = \int_{P_1} f(t)\,dt + \int_{P_2} \frac{F(t)}{2t^\alpha} f(t)\,dt + \int_{P_2} \frac{f(t)}{2}\,dt,$$

where $P_1 = \{t: 0 \leq t \leq n, Z(t) \geq 1\}$ and $P_2 = \{t: 0 \leq t \leq n, 0 \leq Z(t) \leq 1\}$.

By virtue of part iii we know that F is continuous. Hence, we can partition P_1 into k, say, intervals $[p_i, q_i]$ and P_2 into k' intervals $[p_i', q_i']$. Note that although the intervals do not coincide they share the same endpoints—they are the set of zeros in $[0, n]$ of $Z(t) - 1$. Hence,

$$Z(p_i) = Z(q_i) = Z(p_i') = Z(q_i') = 1 \Rightarrow F(p_i) = p_i^\alpha,$$

$$F(q_i) = q_i^\alpha,\ F(p_i') = p_i'^\alpha \text{ and } F(q_i') = q_i'^\alpha.$$

Thus,

$$I_n = \sum_{i=1}^{k}\int_{[p_i,q_i]} f(t)\,dt + \sum_{i=1}^{k'}\int_{[p_i',q_i']}\frac{F(t)}{2t^\alpha} f(t)\,dt + \sum_{i=1}^{k'}\int_{[p_i',q_i']}\frac{f(t)}{2}\,dt$$
$$= \sum_{i=1}^{k}(F(q_i) - F(p_i)) + \frac{1}{2}\sum_{i=1}^{k'}(F(q_i') - F(p_i')) + \sum_{i=1}^{k'}\int_{[p_i',q_i']}\frac{F(t)}{2t^\alpha} f(t)\,dt.$$

If we order the intervals so that $p_{i+1} \geq q_i$ and $p_{i+1}' \geq q_i'$ it is easy to see that

$$\sum_{i=1}^{k}(F(q_i) - F(p_i)) + \sum_{i=1}^{k'}(F(q_i') - F(p_i'))$$

is a telescoping series and so is $0(n^\alpha)$. Thus:

$$I_n = 0(n^\alpha) + \sum_{i=1}^{k'}\int_{p_i'}^{q_i'}\frac{F(t)}{2t^\alpha} f(t)\,dt.$$

Now

$$\int_{p_i'}^{q_i'}\frac{F(t)}{t^\alpha} f(t)\,dt \geq \int_{p_i'}^{q_i'}\frac{F(t)f(t)}{t^\alpha}\,dt - \int_{p_i'}^{q_i'}\frac{\alpha F(t)^2}{2t^{\alpha+1}}\,dt = \frac{F^2(q_i')}{2q_i'^\alpha} - \frac{(F^2p_i')}{2p_i'^\alpha}.$$

Also,

$$\int_{p_i'}^{q_i'}\frac{F(t)}{t^\alpha} f(t)\,dt \leq \int_{p_i'}^{q_i'}\frac{F(t)f(t)}{t^\alpha}\,dt - \int_{p_i'}^{q_i'}\frac{\alpha F(t)^2}{2t^{\alpha+1}}\,dt + \int_{p_i'}^{q_i'}\frac{\alpha t^{\alpha-1}}{2}\,dt = \frac{F^2(q_i')}{2q_i'^\alpha} - \frac{F^2(p_i')}{2p_i'^\alpha} + \frac{q_i'^\alpha - p_i'^\alpha}{2}$$

as $|F(t)| \leq t^\alpha$ for $t \in [p_i', q_i']$. Thus, we may conclude that

$$\int_{p_i'}^{q_i'}\frac{F(t)}{2t^\alpha} f(t)\,dt = \frac{F^2(q_i')}{4q_i'^\alpha} - \frac{F^2(p_i')}{4p_i'^\alpha} + 0(q_i'^\alpha - p_i'^\alpha)$$
$$= \frac{1}{4}(q_i'^\alpha - p_i^\alpha) + 0(q_i'^\alpha - p_i'^\alpha).$$

Hence,

$$I_n = 0(n^\alpha) + \sum_{i=1}^{k'} \left\{\frac{1}{4}(q_i'^\alpha - p_i'^\alpha + 0(q_i'^\alpha - p_i'^\alpha)\right\}$$
$$= 0(n^\alpha).$$

Theorem 1. *If A and B are two forecasters with a_i, b_i are bounded for all i, and C is defined by MM, then C is better than A and B.*

Proof. To prove the theorem it is sufficient to show that for any $\epsilon > 0$ and $\delta > 0$ there is an n sufficiently large subject to

$$\Pr(T(C, n) - (T(A, n), T(B, n))^- > \delta) < \epsilon.$$

From Markov's inequality we deduce that:

$$\Pr(T(C, n) - (T(A, n), T(B, n))^- > \delta)$$
$$\leqslant \frac{E[T(C, n)) - (T(A, n), T(B, n))^-]}{\delta}.$$

So, we need to bound $E[T(C, n) - (T(A, n), T(B, n))^-]$.

Without loss of generality we may assume that $T(A, n) \geqslant T(B, n)$. Let

$$P_i = \left[\left(0, \frac{iD(i) + i^s}{2i^s}\right)^+, 1\right]^-$$

and $F(i) = i\,D(i)$. Let U be a uniform [0, 1] random variable and define $I(x)$ to be 1 if $x \leqslant U$ and zero otherwise. Then,

$$nT(C,n) = nT(B,n) + \sum_{i=0}^{n-1} (a_{i+1} - b_{i+1})I(P_i)$$
$$= nT(B,n) + \sum_{i=0}^{n-1} \{(i+1)D(i+1) - iD(i)\}I(P_i)$$
$$= nT(B,n) + \sum_{i=0}^{n-1} (F(i+1) - F(i))I(P_i).$$

Now, extend $F(t)$ to the reals by setting $F(t)$ equal to $(t - \lfloor t \rfloor)F(\lfloor t \rfloor + 1) + (1 - t + \lfloor t \rfloor)F(\lfloor t \rfloor)$ if t is nonintegral. Then:

$$E(nT(C, n))$$
$$= nT(B, n) + \sum_{i=0}^{n-1} (F(i + 1) - F(i))P_i$$
$$= nT(B, n) + \int_0^n \left[\left(0, \frac{F(\lfloor t \rfloor) + \lfloor t \rfloor^s}{2t^s}\right)^+, 1\right]^- dF(t)$$
$$= nT(B, n) + \int_0^n \left[\left(0, \frac{F(t) + t^s}{2t^s}\right)^+, 1\right]^- dF(t)$$
$$+ 0\left(\int_0^n \frac{dt}{\lfloor t \rfloor^s}\right)$$

as $|F(t) - F(\lfloor t \rfloor)| \leqslant 1$, $|\lfloor t \rfloor - t| \leqslant 1$ and $|dF(t) \leqslant dt$.

Now

$$\int_0^n \frac{dt}{t^s} = 0\,(n^{1-s}).$$

Invoking the lemma above we deduce that

$$E(nT(C, n)) = nT(B, n) + 0(n^s) + 0(n^{1-s})$$
$$E(T(C, n)) = T(B, n) + 0(n^{s-1}) + 0(n^{-s})$$
$$E(T(C, n) - (T(A, n), T(B, n))^-)$$
$$= 0(n^{s-1}) + 0(n^{-s}).$$

Hence,

$$\Pr(T(C, n) - (T(A, n), T(B, n))^- > \delta)$$
$$\leqslant \frac{0(n^{s-1}) + 0(n^{-s})}{\delta}.$$

This proves the theorem.

Theorem 2. $E(\hat{D}(n)|\Sigma) = D(n)$ *and* $\mathrm{Var}(\hat{D}(n)|\Sigma) = 0(n^{-r})$.

Proof. Let

$$m_i = \frac{X_i\,(a_i - b_i)}{n^{r-1}} - (a_i - b_i).$$

Then,

$$\hat{D}(n) = D(n) + \frac{1}{n}\sum_{i=1}^{n} m_i.$$

So,

$$-E(\hat{D}(n)|\Sigma) = D(n) + \frac{1}{n}\sum_{i=1}^{n} E(m_i|\Sigma)$$
$$= D(n) + \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{E(X_i|\Sigma)(a_i - b_i)}{n^{r-1}} - (a_i - b_i)\right\}$$
$$= D(n) + \frac{1}{n}\sum_{i=1}^{n}\{(a_i - b_i) - (a_i - b_i)\}$$
$$= D(n). \text{ As } E(X_i|\Sigma) = E(X_i).$$

From the definition of variance we see that

$$\mathrm{Var}(\hat{D}(n)|\Sigma) = E[(\hat{D}(n) - D(n))^2|\Sigma]$$
$$= E\left[\left(\frac{1}{n}\sum_{i=1}^{n} m_i\right)^2 \middle| \Sigma\right]$$
$$= n^{-2}\sum_{ij} E(m_i m_j|\Sigma)$$
$$= n^{-2}\sum_{i,j} E(m_i^2|\Sigma) \text{ as } E(m_i|\Sigma) = 0$$

and the m_i's are conditionally independent given Σ.

A straightforward calculation shows that $E(m_i^2 \mid \Sigma) \leq (a_i - b_i)^2 n^{1-r}$ and so $\mathrm{Var}(\hat{D}(n) \mid \Sigma) = 0(\mathrm{n}^{-r})$.

Theorem 3. *If A and B are any two forecasters, then $\hat{C}$ is better than A and B.*

Proof. To prove this theorem it will be sufficient to show that $|T(C, n) - T(\hat{C}, n) \rightarrow 0$ in probability as $n \rightarrow \infty$.

From the fact that the forecast errors are bounded it follows that:

$$|C_i - \hat{C}_i| \leq \left| \left[\left(0, \frac{i\,\hat{D}(i) + i^s}{2i^s} \right)^+, 1 \right]^- - \left[\left(0, \frac{i\,D(i) + i^s}{2i^s} \right)^+, 1 \right]^- \right| M$$

$$\leq \frac{1}{2} i^{1-s} |\hat{D}(i) - D(i)| M,$$

where $M = \max_{i \geq 1} \{a_i, b_i\}$.

Thus,

$$E(T(C, n) - T(\hat{C}, n) \mid \Sigma)$$

$$= \frac{1}{n} \sum_{i=1}^{n} E[(C_i - \hat{C}_i) \mid \Sigma]$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} E(|C_i - \hat{C}_i| \mid \Sigma)$$

$$\leq \frac{M}{2n} \sum_{i=1}^{n} i^{1-s} E(|\hat{D}(i) - D(i)| \mid \Sigma)$$

$$\leq \frac{M}{2n} \sum_{i=1}^{n} i^{1-s} [E\{(\hat{D}(i) - D(i))^2 \mid \Sigma\}]^{1/2}$$

$$\leq \frac{M}{2n} \sum_{i=1}^{n} i^{1-s}\, 0(i^{-r/2})$$

$$= 0(n^{1-s-r/2}). \quad \text{(by Theorem 2)}$$

Since this last term is independent of Σ it follows that $E(T(C, n) - T(C, n)) = 0(n^{1-s-r/2})$. As $s + r/2 > 1$ it follows that $0(n^{1-s-r/2}) \rightarrow 0$ as $n \rightarrow \infty$. The result now follows from an application of the Markov inequality.

ACKNOWLEDGMENT

We would like to thank Pat Thompson, Barry Nelson, and Marc Posner for their comments. Also, the exposition and discussion of MM has benefitted greatly from the scrutiny of Robert Clemen and an anonymous referee.

REFERENCES

CLEMEN, R. T., AND R. L. WINKLER. 1986. Combining Economic Forecasts. *J. Bus. and Econ. Statist.* **4,** 39–46.

CLEMEN, R. T. 1989. Combining Forecasts: A Review and Annotated Bibliography. *Intl. J. Forecast.* **5,** 559–583.

GUPTA, S., AND P. WILSON. 1987. Combination of Forecasts: An Extension. *Mgmt. Sci.* **33,** 356–372.

MAKRIDAKIS, S., AND R. L. WINKLER. 1983. Averages of Forecasts: Some Empirical Results. *Mgmt. Sci.* **29,** 987–996.

POPPER, K. 1968. *Conjectures and Refutations.* Routledge and Kegan Paul, London.

SCHNARRS, S. P. 1984. Situational Factors Affecting Forecast Accuracy. *J. Market. Res.* **21,** 290–297.