OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 71, 3 (2009) 0305-9049
doi: 10.1111/j.1468-0084.2008.00541.x

A Simple Explanation of the Forecast Combination Puzzle*

JEREMY SMITH and KENNETH F. WALLIS

Department of Economics, University of Warwick, Coventry CV4 7AL, UK
(e-mail: k.f.wallis@warwick.ac.uk)

Abstract

This article presents a formal explanation of the forecast combination puzzle, that simple combinations of point forecasts are repeatedly found to outperform sophisticated weighted combinations in empirical applications. The explanation lies in the effect of finite-sample error in estimating the combining weights. A small Monte Carlo study and a reappraisal of an empirical study by Stock and Watson [*Federal Reserve Bank of Richmond Economic Quarterly* (2003) Vol. 89/3, pp. 71–90] support this explanation. The Monte Carlo evidence, together with a large-sample approximation to the variance of the combining weight, also supports the popular recommendation to ignore forecast error covariances in estimating the weight.

I. Introduction

The idea that combining different forecasts of the same event might be worthwhile has gained wide acceptance since the seminal article of Bates and Granger (1969). Twenty years later, Clemen (1989) provided a review and annotated bibliography containing over 200 items, which he described as 'an explosion in the number of articles on the combination of forecasts', mostly concerning point forecasts; Wallis (2005) considers extensions to the combination of interval and density forecasts. Despite the explosion of activity, Clemen found that a variety of issues remained to be addressed, the first of which was 'What is the explanation for the robustness of the simple average of

*The first version of this article, entitled 'Combining point forecasts: the simple average rules, OK?', was presented in seminars at Nuffield College, Oxford, and Universidad Carlos III de Madrid in Spring 2005. The helpful comments of seminar participants, anonymous referees, Mark Watson and Kenneth West, and access to the database of James Stock and Mark Watson are gratefully acknowledged.

JEL Classification numbers: C22, C53, E37.

forecasts?' (1989, p. 566). That is, why is it that, in comparisons of combinations of point forecasts based on mean-squared forecast errors (MSFEs), a simple average, with equal weights, often outperforms more complicated weighting schemes. This empirical finding continually reappears, for example, in several articles by Stock and Watson (1999, 2003a, 2004), who call this result the 'forecast combination puzzle' in the most recent of these articles (2004, p. 428). The differences are not necessarily large; for example, in the second article the MSFE improvement of the simple average does not exceed 4% (2003a, Table 4), but they are always in the same direction. A simple explanation of this puzzle is explored in this article.

The explanation rests in part on the observation that, in the more complicated weighting schemes, the weights must be estimated. Bates and Granger (1969) considered the estimation properties of the weight in a simple two-forecast example (albeit assuming, unrealistically, uncorrelated forecast errors), and found that the distribution of the estimated weight is highly dispersed. This finding reappeared in several theoretical and empirical extensions over the following two decades, as noted in Clemen's (1989) survey, and led some authors to speculate that the estimated combining weights might be so unreliable or unstable that the theoretical advantage of the combined forecast over an individual forecast is lost. No attention appears to have been given to the impact on the comparison of different combined forecasts, however, and Clemen's (1989) first question quoted above remained unanswered.

More recently, the effect of parameter estimation error on forecast evaluation procedures has been much studied, following the contribution of West (1996). This analysis is extended to the forecast evaluation of competing nested models by Clark and West (2006). In this article, we observe that the same framework is relevant to the forecast combination puzzle, because more general weighting schemes nest the simple average, and we show how it accommodates a resolution of the puzzle. It is shown that, if the optimal combining weights are equal or close to equality, a simple average of competing forecasts is expected to be more accurate, in terms of MSFE, than a combination based on estimated weights.

The article proceeds as follows. Section II presents the general framework for analysis, beginning with the case of two competing forecasts and then considering generalizations to combinations of many forecasts. Two empirical applications follow: the first, in section III, is a small Monte Carlo study of combinations of two forecasts; the second, in section IV, is a reappraisal of a study by Stock and Watson (2003a) of combinations of many forecasts of US output growth. An appendix to section III develops an asymptotic result that supports the empirical recommendation to ignore forecast error covariances in calculating combining weights. Section V concludes. Forecast comparisons in the cited articles by Stock and Watson, as in many other articles, are based on relatively informal comparisons of performance measures such as MSFE, without formal hypothesis testing, and this article takes the same approach. The article's main contributions are presented in relatively simple settings, to aid communication. For discussion of other aspects of forecast combination in more general settings – asymmetric loss functions, non-stationarities, shrinkage

estimators – see the recent survey by Timmermann (2006). Extensions to formal inference procedures are a subject for further research.

II. Weighted and unweighted combinations of forecasts

2.1. Optimal weights

We follow Bates and Granger (1969) and consider the case of two competing point forecasts, f_{1t} and f_{2t}, made h periods earlier, of the quantity y_t. The forecast errors are

$$e_{it} = y_t - f_{it}, \quad i = 1, 2.$$

It is usually assumed that the forecasts are unconditionally unbiased, or 'unbiased on average' in Granger and Newbold's (1986, p. 144) term, so that

$$E(e_{it}) = 0, \quad i = 1, 2.$$

We denote the forecast error variances as $\sigma_i^2, i = 1, 2$, and their covariance as σ_{12}. The combined forecast is the weighted average

$$f_{Ct} = k f_{1t} + (1 - k) f_{2t}, \tag{1}$$

which is also unbiased in the same sense. Its error variance is minimized by setting the weight k equal to

$$k_o = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}, \tag{2}$$

the 'optimal' value, noting a sign error in Bates and Granger's equation (1). This expression can also be recognized as the coefficient in a regression of e_{2t} on $(e_{2t} - e_{1t})$, which suggests a way of estimating k_o from data on forecasts and outcomes. A further interpretation is that this is equivalent to the extended realization-forecast regression

$$y_t = \alpha + \beta_1 f_{1t} + \beta_2 f_{2t} + u_t \tag{3}$$

subject to the restrictions $\alpha = 0, \beta_1 + \beta_2 = 1$. Although weights outside the (0, 1) interval might be thought to be hard to justify, all these interpretations admit this possibility. An estimate based on equation (2) is negative whenever sample moments satisfy $s_{12} > s_2^2$, and exceeds one if $s_{12} > s_1^2$.

The minimized error variance of the combined forecast is no greater than the smaller of the two individual forecast error variances; hence, in general, there is a gain from combining using the optimal weight. Equality occurs if the smaller variance is that of a forecast which is already the minimum mean-squared error (MMSE) forecast: there is then no gain in combining it with an inferior forecast. If the MMSE forecast is f_{1t}, say, with error variance σ_1^2, then it also holds that $\sigma_{12} = \sigma_1^2$ for any other forecast f_{2t}, whereupon $k_o = 1$. In this case, and if $h = 1$, the error term in equation (3)

is non-autocorrelated. In all other cases, the error term is expected to exhibit autocorrelation; hence, $\hat{k}_o$ or its regression equivalent is not in general fully efficient.

The simple average, with equal weights, is the case $k = 0.5$. This is optimal if $\sigma_1^2 = \sigma_2^2$, i.e. the two competing forecasts are equally good (or bad), irrespective of any covariance between their errors. A further possibility suggested by Bates and Granger is to neglect any covariance term, or assume it to be zero, and use the expression

$$k' = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

This again gives equal weights if the error variances are equal, but also restricts the weights to the $(0, 1)$ interval in general. However, if f_{1t} is the MMSE forecast and f_{2t} is any other forecast, this does not deliver weights of 1 and 0 as in the previous paragraph. That k' is the weight attached to the first forecast, f_{1t}, can be emphasized by expressing it in the alternative form

$$k' = \frac{\dfrac{1}{\sigma_1^2}}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2}}. \tag{4}$$

This makes it clear that the weights are inversely proportional to the corresponding forecast error variances, and gives an expression which is more amenable to generalization below.

2.2. Pseudo out-of-sample comparisons of combined forecasts

The general approach to forecast evaluation that is followed in the literature is called pseudo out-of-sample forecasting by Stock and Watson (2003b, §12.7) in their textbook and empirical studies cited above, because it mimics real-time out-of-sample forecasting yet the 'future' outcomes are known, and so forecast performance can be assessed. To this end, a sample of available data, real or artificial, is divided into two subsamples: the first is used for estimating the forecasting relationships, the second for evaluating their forecast performance. Forecasting with a constant lead time, say one step ahead, implies that the information set on which the forecast is based is updated as the forecast moves through the evaluation subsample, and it is an open question whether and, if so, how the estimated relationships should also be updated. The three possibilities that figure prominently in the literature are referred to as fixed, recursive and rolling schemes, and which scheme is used has a bearing on the asymptotics of the various available tests (Clark and McCracken, 2001). However, many studies of combined forecasts are based on informal comparisons of MSFEs over the evaluation subsample rather than on formal inference procedures, as noted above, whereupon the choice of updating scheme is immaterial. The comparisons developed below remain in this informal mode.

We adopt the Clark–McCracken–West notational conventions and denote, in one-step-ahead forecasting, the size of the available sample as $T+1$. This is divided into the initial estimation ('regression') subsample of R observations and the second evaluation ('prediction') subsample of P observations, with $R+P=T+1$. The first forecast is made at time R of observation y_{R+1}, and the last (the Pth) is made at time T of observation y_{T+1}. The pseudo out-of-sample MSFE of the combined forecast (1) is then

$$\hat{\sigma}^2_{\mathrm{C}} = \frac{1}{P}\sum_{t=R+1}^{T+1}(y_t - f_{\mathrm{C}t})^2.$$

We consider three combined forecasts: the first, denoted f_{st}, is the simple average of forecasts with $k=0.5$ in equation (1), associated forecast error $e_{st}=y_t-f_{st}$ and MSFE $\hat{\sigma}^2_s$; next is the weighted average f_{wt} using an estimate $\hat{k}$ in equation (1), with error e_{wt} and MSFE $\hat{\sigma}^2_w$; finally, and hypothetically, the optimal combined forecast f_{ot} is based on the optimal, but unknown weight k_o and has hypothetical error e_{ot}. Rearranging equation (1) as

$$f_{\mathrm{C}t} = f_{2t} + k(f_{1t} - f_{2t}),$$

specializing to the simple and weighted average combination forecasts in turn, and subtracting the corresponding expression for f_{ot} gives the following relations among the forecast errors:

$$e_{st} - e_{ot} = -(0.5 - k_o)(f_{1t} - f_{2t}), \quad e_{wt} - e_{ot} = -(\hat{k} - k_o)(f_{1t} - f_{2t}).$$

Hence, the typical term in the MSFE difference $\hat{\sigma}^2_s - \hat{\sigma}^2_w$ is

$$e^2_{st} - e^2_{wt} = \{(0.5 - k_o)^2 - (\hat{k} - k_o)^2\}(f_{1t} - f_{2t})^2 + z_t,$$

where the cross-product term $z_t = 2e_{ot}(\hat{k} - 0.5)(f_{1t} - f_{2t})$ has expected value zero, neglecting any correlation between the estimation subsample and the evaluation subsample, and noting that e_{ot} is uncorrelated with $e_{1t} - e_{2t}$. Thus, the MSFE difference is

$$\hat{\sigma}^2_s - \hat{\sigma}^2_w \approx \{(0.5 - k_o)^2 - (\hat{k} - k_o)^2\}\frac{1}{P}\sum_{t=R+1}^{T+1}(f_{1t} - f_{2t})^2, \tag{5}$$

which shows the standard trade-off between bias, from assuming a different value of the 'true' coefficient, and variance, in estimating that coefficient.

In comparing the simple average with the weighted average of forecasts, we note that, as in the framework of Clark and West (2006), the models being compared are nested: the null model has $k=0.5$; under the alternative, $k \neq 0.5$. Putting $k_o = 0.5$ in equation (5) and noting that $f_{wt} - f_{st} = (\hat{k} - 0.5)(f_{1t} - f_{2t})$, under the null we expect to find

$$\hat{\sigma}_s^2 - \hat{\sigma}_w^2 \approx -\frac{1}{P}\sum_{t=R+1}^{T+1}(f_{wt} - f_{st})^2 < 0. \tag{6}$$

Thus, the simple average is expected to outperform the weighted average systematically, in a situation in which they are theoretically equivalent. The 'null discrepancy' on the right-hand side of equation (6) corresponds to the MSFE *adjustment* defined in Clark and West's equation (2.9), and can be calculated directly. It remains of the same order of magnitude as P increases.

We note that this result reverses the choice of combined forecast which might be made using the evidence of the regression subsample. Estimating k gives the weighted average a better fit than the simple average in the estimation subsample, but this amounts to overfitting in the present circumstances, and to choose the weighted average on this basis is the wrong choice for the prediction subsample. Before discussing estimation of k, we consider generalizations to more than two competing forecasts, as the number of forecasts being combined may be relevant to the choice of estimator.

2.3. Combining many forecasts

The general framework presented above readily extends to more than two competing forecasts, although some practical issues arise. With n competing point forecasts $f_{it}, i = 1, \ldots, n,$ the combined forecast is

$$f_{Ct} = \sum_{i=1}^{n} k_i f_{it},$$

with $\sum k_i = 1$ if the individual forecasts are unbiased and this is also desired for the combined forecast. Granger and Ramanathan (1984) consider estimation of the corresponding generalization of regression equation (3), namely

$$y_t = \alpha + \beta_1 f_{1t} + \cdots + \beta_n f_{nt} + u_t, \tag{7}$$

and the question of whether or not the coefficient restrictions $\alpha = 0$ and/or $\sum \beta_i = 1$ should be imposed. The unconstrained regression clearly achieves the smallest error variance *ex post*, and gives an unbiased combined forecast even if individual forecasts are biased. However, if the practical objective is to improve *ex ante* forecast performance, then the imposition of the restrictions improves forecast efficiency, as shown by Clemen (1986), for example.

Estimation of the regression equation (7) runs into difficulty if the number of individual forecasts being combined, n, is close to the number of observations in the regression subsample, R. This is a feature of the applications by Stock and Watson in the three articles referred to in the Introduction. The first article (Stock and Watson, 1999) analyses the performance of 49 linear and nonlinear univariate forecasting methods, in combinations with weights estimated over 60 or 120 months; this

is carried out for 215 different series. The second article (2003a) considers combinations of up to 37 forecasts of US output growth based on individual leading indicators, with weights estimated recursively, with an initial sample of 68 quarterly observations. The third article (2004) extends the growth forecasts to the G7 countries, and the number of individual leading indicators considered for each country ranges between 56 and 75.

In these circumstances, Stock and Watson abandon estimation of the optimal combining weights by regression or, as they put it, abandon estimation of the large number of covariances among the different forecast errors. They follow the suggestion of Bates and Granger (1969) noted in the final paragraph of section 2.1 and base estimated weights on the generalization of expression (4); thus,

$$\hat{k}_i' = \frac{1/s_i^2}{\sum_{j=1}^{n} 1/s_j^2}, \quad i = 1, \ldots, n, \tag{8}$$

where s_i^2, $i = 1, \ldots, n$, is the MSFE of f_{it} over an estimation data set. Earlier empirical studies summarized by Clemen (1989, p. 562) also support the suggestion 'to ignore the effects of correlations in calculating combining weights'. Stock and Watson use several variants of this estimator, of which two are of particular interest. The first is to raise each MSFE term in the above expression to the power ω. With $0 < \omega < 1$, this shrinks the weights towards equality, the case $\omega = 0$ corresponding to the simple average with $k_i = 1/n$. Or with $\omega > 1$, more weight is placed on the better performing forecasts than is indicated by the inverse MSFE weights. The second variant is to calculate MSFEs as discounted sums of past squared forecast errors, so that forecasts that have been performing best most recently receive the greatest weight.

The common finding of the three cited studies in respect of comparisons of different combined forecasts is described, as noted above, as the 'forecast combination puzzle – the repeated finding that simple combination forecasts outperform sophisticated adaptive combination methods in empirical applications' (Stock and Watson, 2004, p. 428). The differences are not necessarily large; for example, in the second article the MSFE improvement of the simple average does not exceed 4% (2003a, Table 4), but there are no reversals.

The explanation advanced in section 2.2 carries over to the case of $n > 2$ competing forecasts. Under the null, the simple average is expected to outperform the weighted average in terms of their MSFEs, in the absence of the adjustment defined in equation (6). Given $n \times 1$ vectors of forecasts and estimated weights, the weighted combination forecast is $f_t'\hat{k}$. Taking expectations in the R sample, and conditioning on the P sample, the expected difference in MSFE is equal to trace $(\Sigma_{ff} V_{\hat{k}})$, where Σ_{ff} is the P-sample moment matrix of the forecasts and $V_{\hat{k}}$ is the mean-squared error matrix of $\hat{k}$ around the true value of equal weights. Thus, the expected discrepancy is smaller, the more accurate, in this sense, are the estimates of the weights. This expression simplifies to the null discrepancy or MSFE adjustment defined in equation (6).

III. Empirical application: a Monte Carlo study

3.1. Experimental design

Our first experiment describes the behaviour of the MSFE discrepancy $\hat{\sigma}_s^2 - \hat{\sigma}_w^2$ analysed in section 2.2 under the null that $k = 0.5$. Accordingly, we construct the two competing forecasts with equal error variances, and compare their simple average to a weighted average with an estimated weight. The data-generating process is the Gaussian autoregressive AR(2) process

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2),$$

subject to the stationarity conditions $\phi_2 < 1 + \phi_1, \phi_2 < 1 - \phi_1, -1 < \phi_2 < 1$. The first two autocorrelation coefficients are then

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}, \quad \rho_2 = \phi_1 \rho_1 + \phi_2.$$

We set up two cases of competing one-step-ahead forecasts with equal error variances. In each case, the competing forecasts are mis-specified; the MMSE forecast based on the AR(2) model does not feature in our comparisons.

Case 1. In the first case, both forecasts are based only on the most recent observation. The first forecast is the naïve 'no-change' forecast and the second forecast is a first-order autoregression with the same forecast error variance. Thus,

$$f_{1t} = y_{t-1}, \quad f_{2t} = (2\rho_1 - 1) y_{t-1}; \quad \sigma_i^2 = 2(1 - \rho_1)\sigma_y^2, \ \ i = 1, 2.$$

The contemporaneous correlation between the two forecast errors is equal to ρ_1. We consider values of ϕ_1 of 0.4 and 0.8, with ϕ_2 taking values in the range $-1 < \phi_2 < 1 - \phi_1$; hence, the forecast errors are positively correlated, with ρ_1 lying in the range $0.2 < \rho_1 < 1$ or $0.4 < \rho_1 < 1$ respectively. In our experiments, the ϕ_2 values are varied by steps of 0.1, except that the non-stationary boundaries are avoided by taking a minimum value of -0.98 and a maximum value of 0.58 or 0.18 respectively.

Case 2. In the second case, the two forecasts are again based on only a single observation, but now with either a one-period or a two-period lag; each forecast is unbiased, conditional on its limited information set. Thus,

$$f_{1t} = \rho_1 y_{t-1}, \quad f_{2t} = \rho_2 y_{t-2}; \quad \sigma_i^2 = (1 - \rho_i^2)\sigma_y^2, \ \ i = 1, 2.$$

To equate the error variances, we choose parameter values such that $\rho_1^2 = \rho_2^2$, specifically $\phi_1 = \phi_2$ to deliver $\rho_1 = \rho_2$, or $\phi_1 = -\phi_2$ to deliver $\rho_1 = -\rho_2$. With these restrictions, stationarity requires that $-1 < \phi_2 < 0.5$, with $\phi_1 = \pm\phi_2$ as appropriate. If $\phi_1 = \phi_2 = 0$, there is an obvious singularity: the series is white noise, the forecasts are equal to the mean of zero and have the same error, and k_o is indeterminate.

In each case, 1,000 artificial time-series samples are generated for each parameter combination. After discarding a 'start-up' portion, each sample is divided into an estimation subsample of R observations and an evaluation subsample of P observations. The forecast parameter values are assumed known, and the estimation subsample, kept fixed and not updated, is used simply to estimate the combining weights k_o and k', by replacing the theoretical moments in equations (2) and (4), respectively, by their sample equivalents. Estimates based on equation (2) need not satisfy $0 \leq \hat{k}_o \leq 1$, as noted in section 2.1, especially if the correlation between the competing forecast errors is large and positive, and we consider two possibilities. One is to use the observed point estimate; the second, following widespread practice endorsed by Granger and Newbold (1986, §9.2), is to replace an estimate outside this range by the nearest boundary value, 0 or 1 as appropriate. There are then four combined forecasts whose MSFEs are calculated over the last P observations: three weighted averages, in turn using $\hat{k}'$ and the observed and truncated $\hat{k}_o$, and the simple average using $k = 0.5$. The estimation cost of each weighted average is expressed as the percentage increase in MSFE above that of the simple average, namely $100(\hat{\sigma}_w^2 - \hat{\sigma}_s^2)/\hat{\sigma}_s^2$.

3.2. Results for case 1

Figure 1 shows the mean (over 1,000 replications) percentage increase in MSFE over the simple average for the three combined forecasts based on estimated weights, with subsample sizes $R = 30$ and $P = 6$. The cost of estimating k is in general positive, as anticipated. However, the different estimates give substantially different performances, for both values of ϕ_1 used in these examples. First, estimating the optimal weight, including the covariance term, increases the MSFE of the weighted average by a rather greater amount than when using the estimate that neglects the forecast error correlation. Second, restricting the point estimate of the optimal weight to the (0, 1) interval makes little difference when the correlation between the forecast errors is low, but improves the performance of the combined forecast as this correlation increases. The forecast error correlation increases with ϕ_2, and is equal to 0.4 at the point at which the two upper plots begin to diverge in panel (a) of Figure 1; its value is 0.57 at the equivalent point in panel (b). In each panel, the last points plotted refer to an almost degenerate case in which the first-order autocorrelation coefficient of the data is close to 1 and the two competing forecasts, and hence their errors, are close to equality.

The average cost of estimating the weight scarcely exceeds 5% for any of the parameter combinations considered, and is often somewhat smaller. However, the sampling variation in the MSFE cost is also small, such that the frequency with which the simple average beats the weighted average across 1,000 replications is typically high. Whenever the mean MSFE cost plotted in Figure 1 exceeds 0.5%, the proportion of times the simple average dominates exceeds 90%; this proportion increases as the mean cost increases.

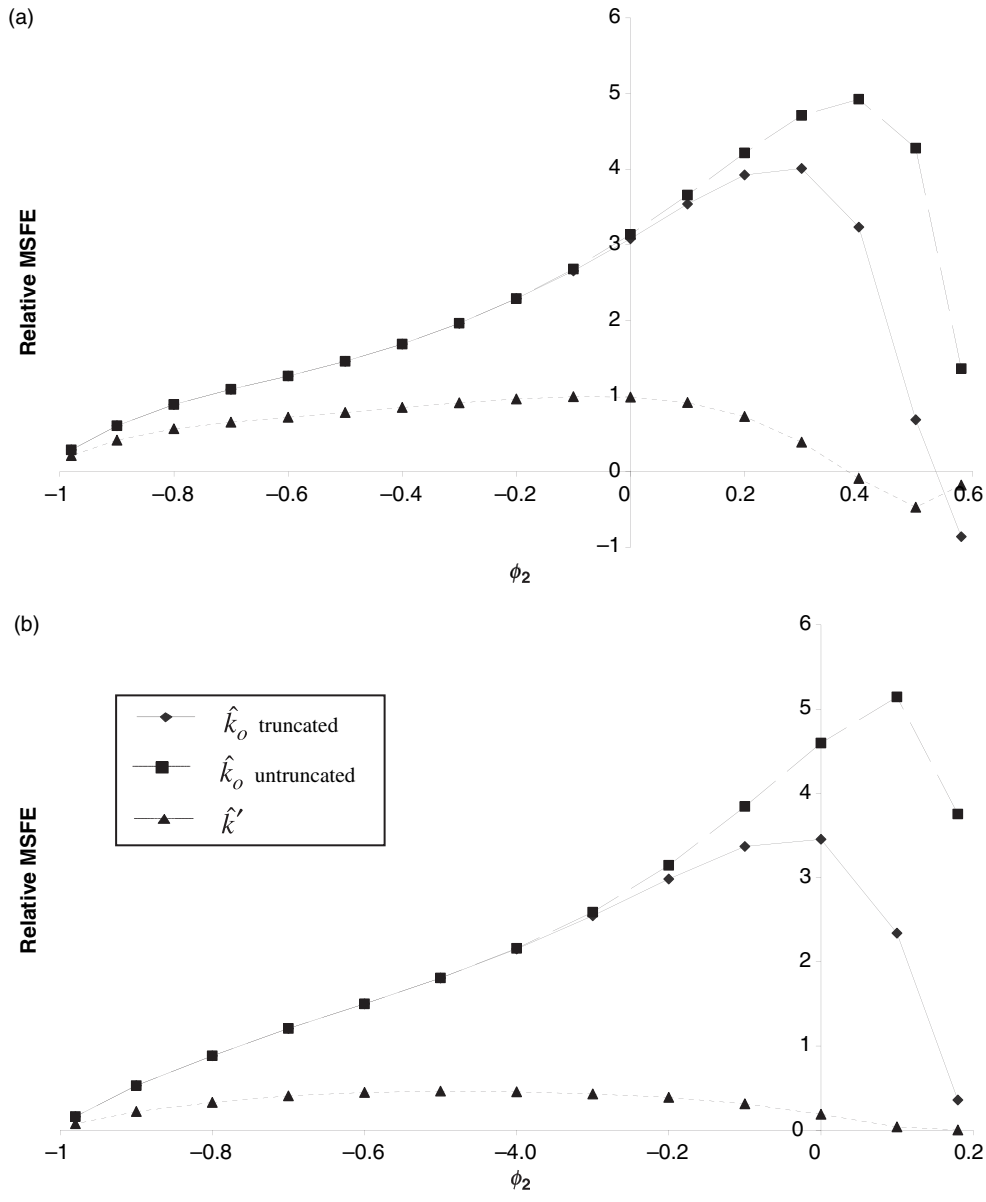

Figure 1. Percentage MSFE cost of weighted combination forecasts; case 1. (a) $\phi_1 = 0.4, R = 30, P = 6$; (b) $\phi_1 = 0.8, R = 30, P = 6$

The key to the differences shown in Figure 1 is the sampling distribution of the three different estimates of the weight. These are shown in Figure 2 for a parameter combination at which the two upper lines in panel (b) of Figure 1 are clearly separated, but not extremely so: the values are $\phi_1 = 0.8, \phi_2 = -0.1$, at which the forecast error correlation, neglected in the k' formula, is equal to 0.73.

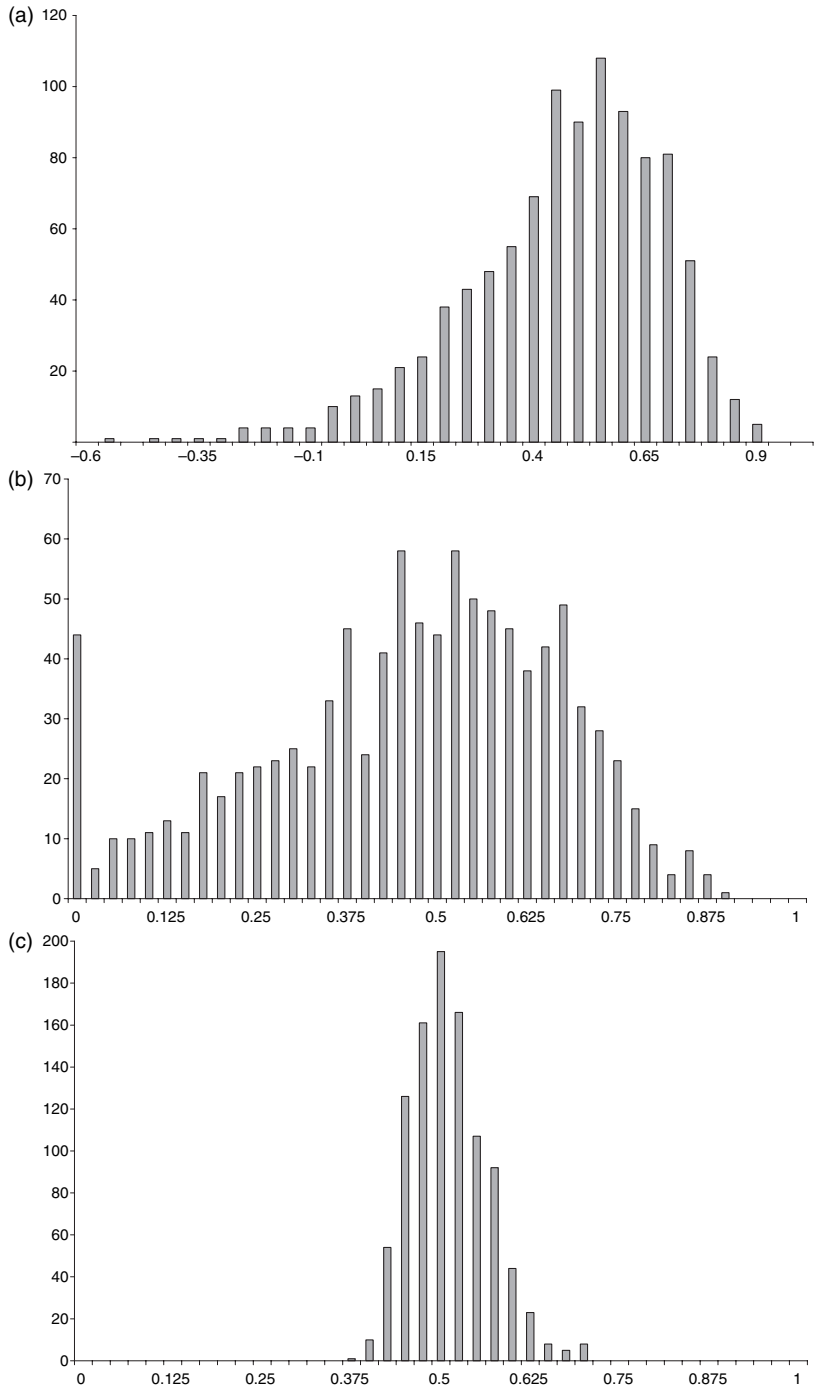

Figure 2. Histograms of estimated weights; case 1, $\phi_1 = 0.8, \phi_2 = -0.1, R = 30$. (a) $\hat{k}_o$, untruncated; (b) $\hat{k}_o$, truncated; (c) $\hat{k}'$

The distribution of the initial estimate of the optimal weight is shown in panel (a) of Figure 2. In our sample of 1,000, there are 44 negative estimates, and these are set equal to zero in panel (b), which results in an improvement in relative MSFE of some 0.5%, as shown in Figure 1b. Note the changes in scale between the panels of Figure 2: in particular, panels (b) and (c) have the same horizontal scale, to emphasize the much smaller dispersion of the estimate $\hat{k}'$, which in turn calls for a change in vertical scale. The better performance of $\hat{k}'$ in this experiment is striking, and supports the recommendation to ignore the error covariances noted above, based on practical applications.

Further support for a preference for $\hat{k}'$ over $\hat{k}_o$ is provided by the asymptotic approximations to the variances of these estimators calculated in the Appendix. These obey the relation

$$\text{asy var}(\hat{k}') = (1-\rho)^2 \,\text{asy var}(\hat{k}_o)$$

where ρ is the forecast error correlation coefficient. This correlation is positive in our experiments, and can be expected to be positive more generally, as the innovation ε_t is common to the competing forecast errors. The formulae developed in the Appendix are seen to provide a good approximation to the simulation variance of the estimates obtained from samples of $R = 30$ observations. More generally, these results offer an explanation of the relatively poor performance of combinations based on the optimal weight.

3.3. Results for case 2

The results shown in Figure 3 are qualitatively similar to those reported for case 1. The cost of estimating k is in general positive. Again the different estimates of the weights yield substantially different performances, the ranking of the different estimates remaining as seen in case 1, with the performance of $\hat{k}'$ again markedly superior to that of $\hat{k}_o$, thanks to the much smaller dispersion of its sampling distribution. Comparing the performance of the two estimates of k_o, Figure 3a shows that at $\phi_1 = \phi_2 = -0.5$, at which the forecast error correlation, neglected in the k' expression, is 0.833, truncating the original estimate gives an improvement in relative MSFE of 0.4%: this is the result of setting 39 negative estimates equal to zero and 34 estimates equal to one, in our sample of 1,000. For $\phi_1 = -\phi_2 = 0.5$ (see Figure 3b) there is an improvement in relative MSFE of 0.35%; here the error correlation is slightly smaller, at 0.805, and slightly fewer $\hat{k}_o$ values are set to the boundary values, 35 and 23 respectively. An example of the sampling distributions of the weights presented in Figure 4 shows the truncation effects diagrammatically, also that $\hat{k}'$ again has much smaller dispersion. In both panels of Figure 3, the truncation effect increases as ϕ_1 approaches zero from above or below, when the correlation between the forecast errors increases towards 1 and we approach the singularity at $\phi_1 = \phi_2 = 0$ noted above.

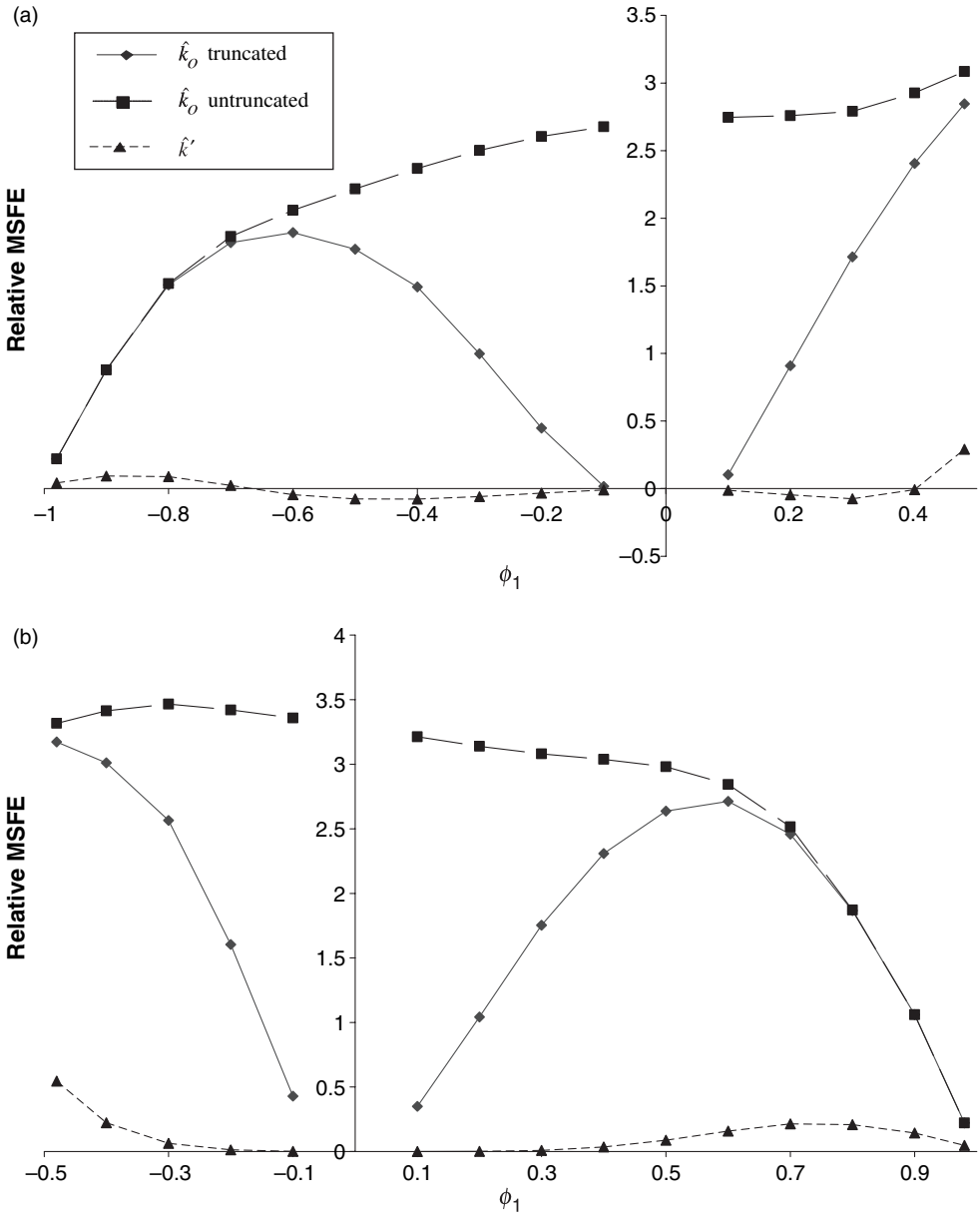

Figure 3. Percentage MSFE cost of weighted combination forecasts; case 2. (a) $\phi_1 = \phi_2, R = 30, P = 6$; (b) $\phi_1 = -\phi_2, R = 30, P = 6$

The behaviour of the estimates of the optimal weight differs between the examples of case 1 and case 2 discussed above. In the first case, truncation of the initial estimate is necessary on only one side; thus, in Figure 2b there is a pile-up at zero but not at one. In the second case, the $(0, 1)$ interval is breached on both

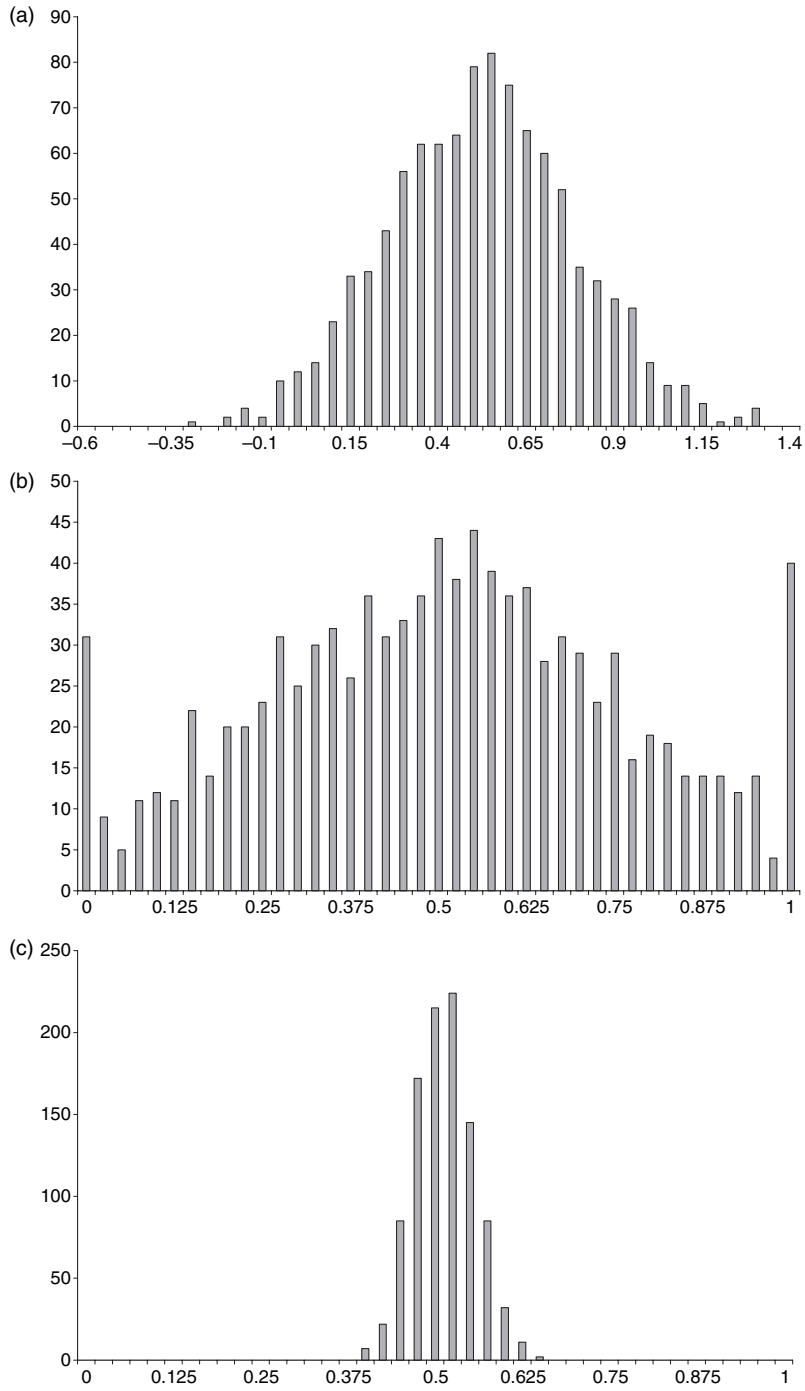

Figure 4. Histograms of estimated weights; case 2, $\phi_1 = -\phi_2 = 0.5, R = 30$. (a) $\hat{k}_o$, untruncated; (b) $\hat{k}_o$, truncated; (c) $\hat{k}'$

sides, as shown in Figure 4. Recalling that $\hat{k}_o < 0$ if $s_2^2 < s_{12}$ and $\hat{k}_o > 1$ if $s_1^2 < s_{12}$, and that the experiment is designed with $\sigma_1^2 = \sigma_2^2$, the explanation of the asymmetry lies in the sampling distribution of the variance estimates. In the example of case 1 shown in Figure 2, the sample variance of s_2^2 is some 35% greater than that of s_1^2, resulting in a tail of the distribution such that $s_2^2 < s_{12}$ in 44 of 1,000 occasions, whereas the equivalent condition for $\hat{k}_o > 1$ never occurs in this sample.

The effects under discussion are finite-sample estimation effects, relating to the size of the 'regression' sample, R, not the 'prediction' sample, P. Increasing P in our experiments has little effect on the mean of the MSFE costs, such as those plotted in Figures 1 and 3, although their sampling variance falls, as expected. Increasing R, however, reduces the MSFE cost of the weighted average, due to increased accuracy of the estimated weight, and at $R = 1,200$ there is essentially no gain in using the simple average, and hence no puzzle.

3.4. Departures from equal weights

We briefly consider the alternative hypothesis $k_o \neq 0.5$, and the trade-off between bias and variance described in equation (5). A relevant question is how different must the optimal weights be for the bias effect from assuming them equal to exceed the estimation variance effect, so that the combination with estimated weights beats the simple average? To admit a non-zero bias in the simple average, we alter one of the competing forecasts considered above so that their error variances are no longer equal. For particular parameter combinations of case 1 and case 2, respectively, we change the coefficient on the lagged observation used to construct the second forecast, f_{2t}, in each case, in order to change its forecast error variance σ_2^2 away from equality with σ_1^2, and hence to change the required combining weight. This weight is indicated by the associated value of k' given by equation (4), deleting the forecast error covariance term in the light of the above results. Otherwise the experimental design remains unaltered, 1,000 replications being undertaken to estimate the average percentage MSFE cost at each of a range of values of k'. When this is equal to 0.5 the results correspond to those given above; departures from 0.5 are indicated by referring to the experiments as case 1* and case 2* respectively.

The results are presented in Figures 5 and 6. When the average percentage MSFE cost takes a negative value, the combination with an estimated weight is doing better than the simple average. The results might appear to show that, in these examples, 'large' departures from equal weights are not required in order to turn the comparison around in this way. The MSFE cost of the weighted estimate using $\hat{k}_o$ is greater than that using $\hat{k}'$; hence, positive costs persist for greater departures from equality. However, in all the cases presented, the cost has turned negative, i.e. the simple average has lost its advantage, before the weights are as different as (0.4, 0.6). In numerical terms this does not seem to be a large departure from equality, but this combination only becomes appropriate if one forecast has MSFE 50% greater than its competitor. Differences as great as this in the performance of competing

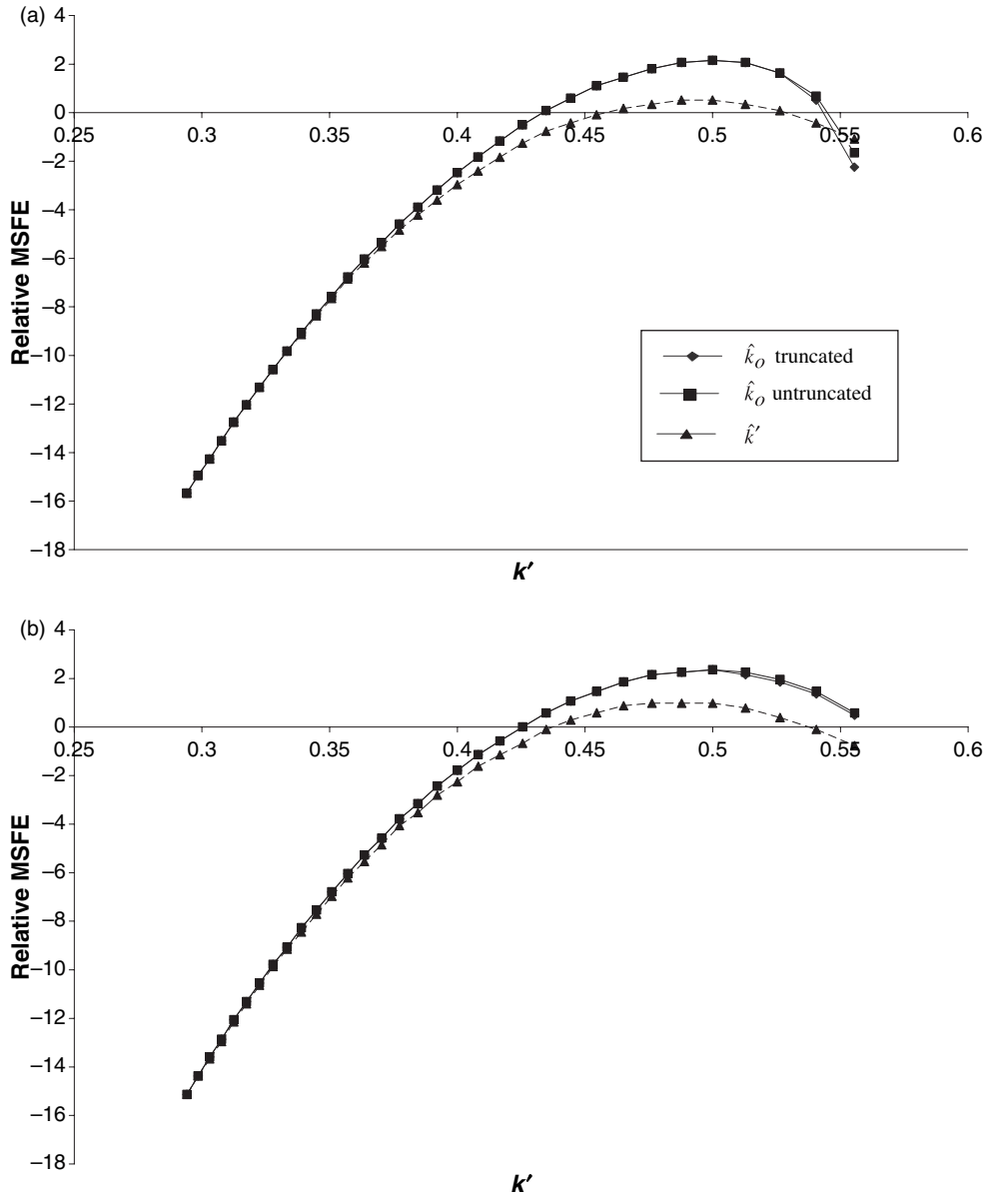

Figure 5. Percentage MSFE cost of weighted combination forecasts; case 1*. (a) $\phi_1 = 0.8, \phi_2 = -0.4$, $R = 30, P = 6$; (b) $\phi_1 = 0.4, \phi_2 = -0.2, R = 30, P = 6$

forecasts are relatively unusual in empirical studies such as those of Stock and Watson cited above, and so the bias effect does not dominate, hence the puzzle. We return to this question following our second empirical illustration, taken from one of their studies.

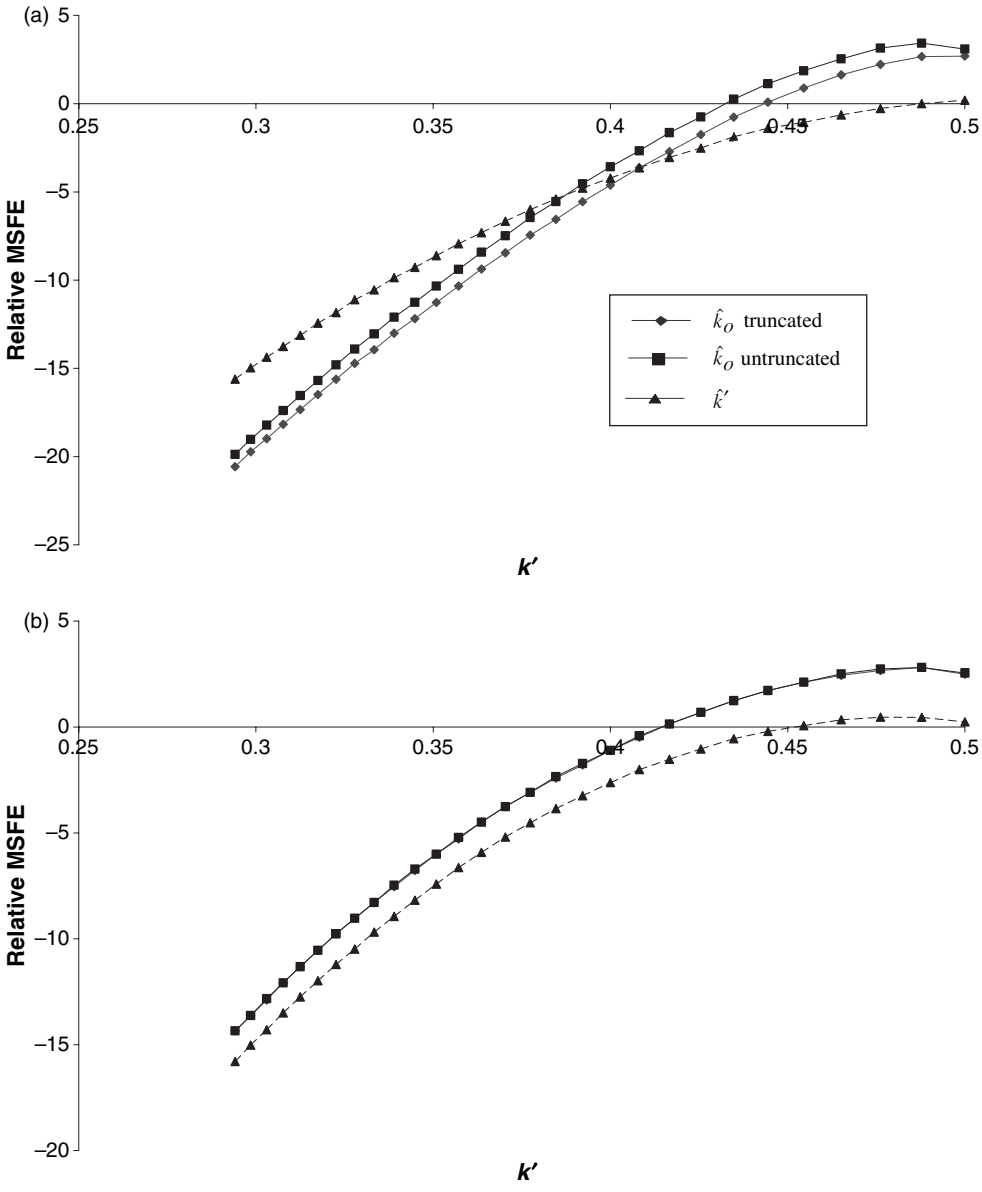

Figure 6. Percentage MSFE cost of weighted combination forecasts; case 2*. (a) $\phi_1 = \phi_2 = 0.45$, $R = 30, P = 6$; (b) $\phi_1 = -\phi_2 = 0.7, R = 30, P = 6$

IV. Empirical application: forecasts of US output growth during the 2001 recession

For a practical application, we revisit the study of Stock and Watson (2003a), which evaluates the performance of leading indicator forecasts during the 2001 recession

TABLE 1

MSFEs of combined quarterly forecasts of output growth (annual percentage rates)

	RGDP		*IP*	
	$h=2$	$h=4$	$h=2$	$h=4$
$\hat{\sigma}_s^2$	1.0078	3.8730	4.4663	23.0034
$\hat{\sigma}_w^2$	1.0319	4.0194	4.4926	23.2114
$\hat{\sigma}_s^2 - \hat{\sigma}_w^2$	−0.0241	−0.1464	−0.0263	−0.2080
Null discrepancy	−0.0005	−0.0029	−0.0032	−0.0082
Remainder	−0.0236	−0.1435	−0.0231	−0.1998

in the USA. This recession differed in many ways from its predecessors, and Stock and Watson find that individual leading indicators also performed differently before and during this recession. Their results show that there were gains to be had from combining forecasts.

Of particular interest for our present purpose is their Table 4, which reports relative MSFEs of various combination forecasts of annual growth rates of real GDP (RGDP) and the Index of Industrial Production (IP) over the period 1999Q1–2002Q3, which spans the recession. Forecast lead times of $h=2$ and 4 quarters are considered; thus, $P=13$ when $h=2$ and $P=11$ when $h=4$. As in all the examples cited in section 2.3, the simple average of $n=35$ competing forecasts, each based on an individual leading indicator, does better than a weighted average using inverse MSFE weights as in equation (8). The weights are based on MSFEs calculated as discounted sums of past squared forecast errors, with a quarterly discount factor of 0.95, over the period from 1982Q1 to the forecast origin, hence R ranges between 65 and 79. From their programs and database we recreate the combined forecast MSFEs on which the relative MSFEs reported in their table are based. (Throughout their article Stock and Watson report the MSFEs of individual and combined forecasts relative to the MSFE of a benchmark autoregressive forecast, whereas we need the numerators of these ratios.) We also calculate the null discrepancy or MSFE adjustment defined in equation (6). The results are shown in Table 1.

The first two rows of the table give the MSFEs of the simple and weighted average forecasts, and the MSFE difference is given in the third row. It is seen that the simple average is the better forecast, by a small margin, in all four cases. The null discrepancy defined in equation (6) is reported in the fourth row, and this is seen to represent only a small part of the MSFE difference. The remainder, which has expected value zero under the equal-weight null hypothesis, clearly has a non-zero sample value. The expression in equation (6) gives the component in the fourth row as the (negative of the) average squared difference between the two combination forecasts or, equivalently, between their forecast errors, and its small size relative to the separate MSFEs suggests that these forecasts are very close to one another. To check this, we plot the forecast errors for all four cases under consideration in Figure 7, which confirms this impression. The two combination forecast errors are virtually indistinguishable,

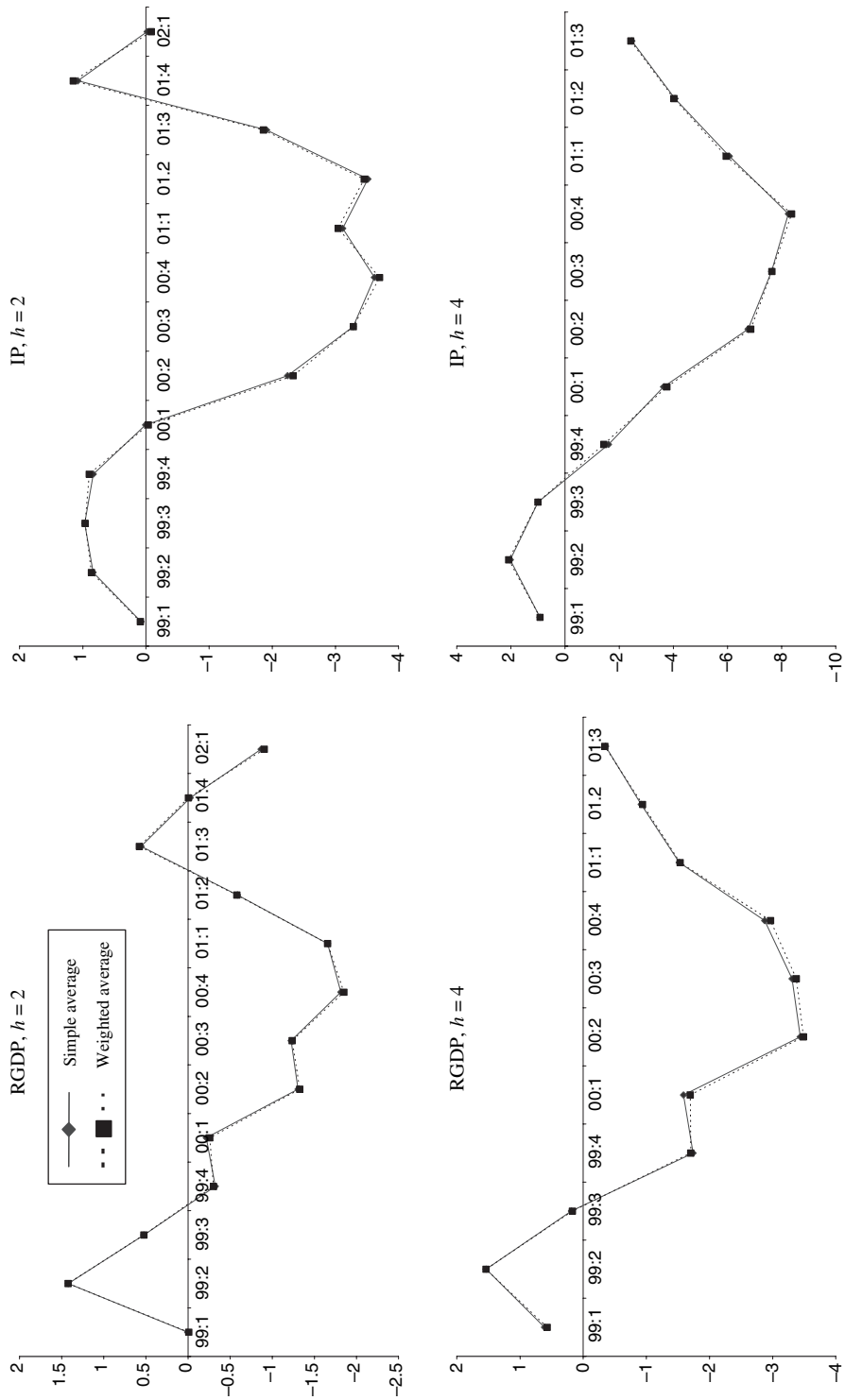

Figure 7. Errors of combined quarterly forecasts of output growth (annual percentage rates)

and the difference between the two combination forecasts is unlikely to be of any importance in the context of a practical decision problem.

The dates shown in Figure 7 correspond to the date of the forecast, the final outcome available being that for 2002Q3. Stock and Watson note that at the time of writing the NBER had not yet dated the trough: on 17 July 2003 this was announced as November 2001 which, with the previous peak at March 2001, gave a contraction duration of 8 months. Figure 7 shows that the combination forecasts substantially overestimated growth throughout this period, starting from the quarter before the peak and, in the year-ahead forecasts, extending well into the recovery phase.

In this example, the forecast combination puzzle is of no importance from a practical point of view. From a statistical point of view, it is an example of the gain in efficiency that can be obtained by imposing, rather than estimating, a restriction that is approximately true. The distribution of estimated weights at the start of the prediction period for the example in column 1 of Table 1 is presented in Figure 8, and this shows rather little variation around the value of $1/n = 0.029$ used by the simple average. The performance of the individual indicators varies over time, hence so do their optimal combining weights, but when the relative weights are small this variation is also likely to have little practical significance.

Optimal forecast weights are close to equality whenever the MSFEs of the individual component forecasts are close to equality. Data on the individual indicator forecast performance presented by Stock and Watson (2003a, Table 3) accord with the impression given by Figure 8: there are some outlying good and bad forecasts, such that the MSFE of the worst forecast is approximately two-and-a-half times that

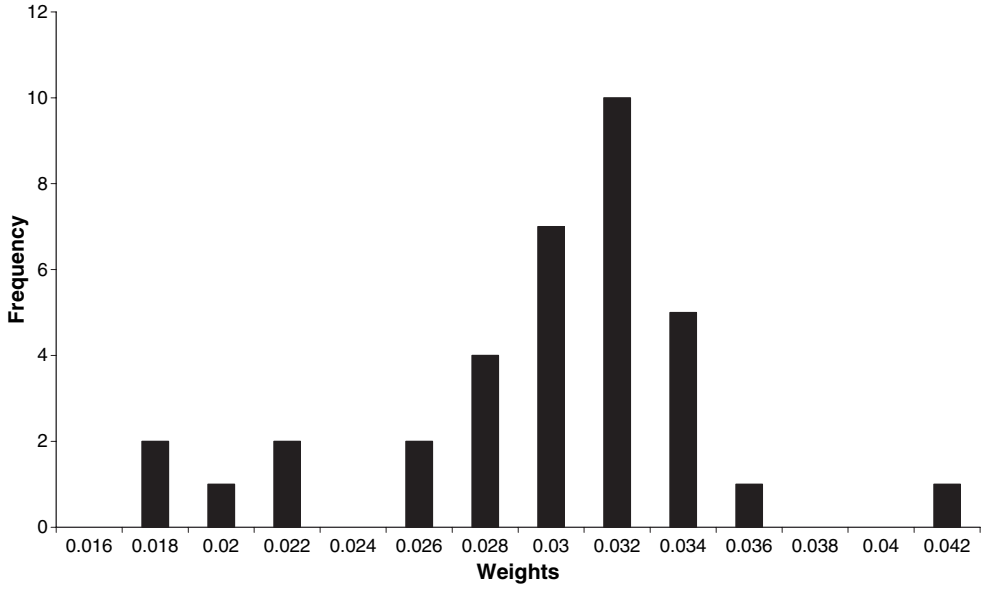

Figure 8. Distribution of weights in weighted average RGDP forecast ($n = 35, h = 2$, first period)

of the best, but the majority differ much less. Each forecast is based on an h-step-ahead distributed lag model expressing Y_{t+h} in terms of past values of the dependent variable, $Y_t, Y_{t-1}, \ldots$, and an individual indicator variable $X_t, X_{t-1}, \ldots$: the separate indicator variables in general vary little in the additional information they bring to the model. Writing the optimal forecast under squared error loss as the conditional expectation given the relevant information set

$$f_{o,t+h} = E(Y_{t+h} \mid \Omega_t),$$

also defines the forecast error $u_{o,t+h} = Y_{t+h} - f_{o,t+h}$ as the cumulated effect of innovations to the Y-process over $t+1, \ldots, t+h$. Any other forecast $f_{i,t+h}(\Omega_{it})$ based on different information, different functional forms, and so forth shares this forecast error: its error $u_{i,t+h}$ is given as

$$\begin{aligned} u_{i,t+h} &= Y_{t+h} - f_{i,t+h}(\Omega_{it}) = u_{o,t+h} + \{E(Y_{t+h} \mid \Omega_t) - f_{i,t+h}(\Omega_{it})\} \\ &= u_{o,t+h} + \delta_{i,t+h}, \text{ say.} \end{aligned}$$

Thus, the underlying innovations provide a floor to the measures of forecast performance that masks the comparative effect of the individual forecast discrepancies $\delta_{i,t+h}$. This appears to have been especially so during the 2001 recession.

Analyses of surveys of economic forecasters often find greater heterogeneity in individual forecast performance than is observed in the above example. Individual judgement plays a greater role in forecast surveys than in comparisons of forecasts from competing statistical models. For the US Survey of Professional Forecasters (SPF) see Davies and Lahiri (1999), and for the Bank of England Survey of External Forecasters see Boero, Smith and Wallis (2008), for example. The survey proprietors – see also *Consensus Economics* – nevertheless present simple averages across survey respondents in their forecast publications (the SPF also reports median responses), rather than weighted averages based on past forecast performance. On the other hand, they are also interested in reporting current sentiment, good or bad, about future macroeconomic developments.

V. Conclusions

Three main conclusions emerge from the foregoing analysis.

If the optimal combining weights are equal or close to equality, a simple average of competing forecasts is expected to be more accurate, in terms of MSFE, than a combination based on estimated weights. The parameter estimation effect is not large, nevertheless it explains the forecast combination puzzle.

However, if estimated weights are to be used, then it is better to neglect any covariances between forecast errors and base the estimates on inverse MSFEs alone, than to use the optimal formula originally given by Bates and Granger for two forecasts, or its regression generalization for many forecasts. This is a familiar recommendation

in the literature, based on empirical studies. Our asymptotic approximation to the variance of the estimated weight provides it with a firmer foundation.

When the number of competing forecasts is large, so that under equal weighting each has a very small weight, the simple average can gain in efficiency by trading off a small bias against a larger estimation variance. Nevertheless, in an example from Stock and Watson (2003a), we find that the forecast combination puzzle rests on a gain in MSFE that has no practical significance.

Final Manuscript Received: October 2008

References

Bartlett, M. S. (1946). 'On the theoretical specification and sampling properties of autocorrelated time-series', *Journal of the Royal Statistical Society Supplement*, Vol. 8, pp. 27–41.

Bartlett, M. S. (1955). *An Introduction to Stochastic Processes with Special Reference to Methods and Applications*, Cambridge University Press, Cambridge.

Bates, J. M. and Granger, C. W. J. (1969). 'The combination of forecasts', *Operational Research Quarterly*, Vol. 20, pp. 451–468.

Boero, G., Smith, J. and Wallis, K. F. (2008). 'Evaluating a three-dimensional panel of point forecasts: the Bank of England Survey of External Forecasters', *International Journal of Forecasting*, Vol. 24, pp. 354–367.

Clark, T. E. and McCracken, M. W. (2001). 'Tests of equal forecast accuracy and encompassing for nested models', *Journal of Econometrics*, Vol. 105, pp. 85–110.

Clark, T. E. and West, K. D. (2006). 'Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis', *Journal of Econometrics*, Vol. 135, pp. 155–186.

Clemen, R. T. (1986). 'Linear constraints and the efficiency of combined forecasts', *Journal of Forecasting*, Vol. 5, pp. 31–38.

Clemen, R. T. (1989). 'Combining forecasts: a review and annotated bibliography', *International Journal of Forecasting*, Vol. 5, pp. 559–583.

Davies, A. and Lahiri, K. (1999). 'Re-examining the rational expectations hypothesis using panel data on multi-period forecasts', in Hsiao C., Pesaran M. H., Lahiri K. and Lee L. F. (eds), *Analysis of Panels and Limited Dependent Variable Models*, Cambridge University Press, Cambridge, pp. 226–254.

Granger, C. W. J. and Newbold, P. (1986). *Forecasting Economic Time Series*, 2nd edn, Academic Press, London.

Granger, C. W. J. and Ramanathan, R. (1984). 'Improved methods of combining forecasts', *Journal of Forecasting*, Vol. 3, pp. 197–204.

Stock, J. H. and Watson, M. W. (1999). 'A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series', in Engle R. F. and White H. (eds), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press, Oxford, pp. 1–44.

Stock, J. H. and Watson, M. W. (2003a). 'How did leading indicator forecasts perform during the 2001 recession?' *Federal Reserve Bank of Richmond Economic Quarterly*, Vol. 89/3, pp. 71–90.

Stock, J. H. and Watson, M. W. (2003b). *Introduction to Econometrics*, Addison Wesley, Boston, MA.

Stock, J. H. and Watson, M. W. (2004). 'Combination forecasts of output growth in a seven-country data set', *Journal of Forecasting*, Vol. 23, pp. 405–430.

Stuart, A. and Ord, J. K. (1994). *Kendall's Advanced Theory of Statistics*, 6th edn, Vol. 1, Edward Arnold, London.

Timmermann, A. (2006). 'Forecast combinations', in Elliott G., Granger C. W. J. and Timmermann A. (eds), *Handbook of Economic Forecasting*, North-Holland, Amsterdam, pp. 135–196.
Wallis, K. F. (2005). 'Combining density and interval forecasts: a modest proposal', *Oxford Bulletin of Economics and Statistics*, Vol. 67, pp. 983–994.
West, K. D. (1996). 'Asymptotic inference about predictive ability', *Econometrica*, Vol. 64, pp. 1067–1084.

Appendix: The estimation variance of the combining weights

We calculate large-sample approximations to the variances of the two estimators of the combining weight, in the case of two forecasts with equal error variances $\sigma_1^2 = \sigma_2^2 = \sigma_e^2$, say. The 'inverse MSFE' coefficient based on equation (4), which neglects the covariance between the forecast errors, is

$$\hat{k}' = \frac{(1/R)\,\Sigma e_{2t}^2}{(1/R)\,\Sigma e_{1t}^2 + (1/R)\,\Sigma e_{2t}^2} = \frac{s_2^2}{s_1^2 + s_2^2} = \frac{1}{1 + (s_1^2/s_2^2)},$$

where R again denotes the estimation sample size. We use standard results on the variance of functions of random variables obtained via Taylor series approximations, sometimes called the 'delta method'; see Stuart and Ord (1994, §10.6), for example. For the nonlinear transformation, we have

$$\mathrm{var}(1+x)^{-1} \approx \left(\frac{\partial(1+x)^{-1}}{\partial x}\right)^2 \mathrm{var}(x),$$

and evaluating the derivative at the mean of 1 gives

$$\mathrm{var}(\hat{k}') \approx \frac{1}{16}\mathrm{var}\left(\frac{s_1^2}{s_2^2}\right).$$

Using the expression for the variance of a ratio of positive random variables, we then have

$$\begin{aligned}\mathrm{var}(\hat{k}') &\approx \frac{1}{16}\left(\frac{E(s_1^2)}{E(s_2^2)}\right)^2 \left(\frac{\mathrm{var}(s_1^2)}{E^2(s_1^2)} + \frac{\mathrm{var}(s_2^2)}{E^2(s_2^2)} - \frac{2\mathrm{cov}(s_1^2, s_2^2)}{E(s_1^2)E(s_2^2)}\right) \\ &= \frac{1}{16\sigma_e^4}[\mathrm{var}(s_1^2) + \mathrm{var}(s_2^2) - 2\ \mathrm{cov}(s_1^2, s_2^2)]. \qquad \text{(A.1)}\end{aligned}$$

Turning to the optimal weight based on equation (2), the estimate is

$$\hat{k}_o = \frac{(1/R)\,\Sigma e_{2t}^2 - (1/R)\,\Sigma e_{1t}e_{2t}}{(1/R)\,\Sigma e_{1t}^2 + (1/R)\,\Sigma e_{2t}^2 - (2/R)\,\Sigma e_{1t}e_{2t}} = \frac{s_2^2 - s_{12}}{s_1^2 + s_2^2 - 2s_{12}} = \frac{1}{1 + \dfrac{s_1^2 - s_{12}}{s_2^2 - s_{12}}}.$$

This last expression is of the same form as $\hat{k}'$, with $s_i^2 - s_{12}$ replacing $s_i^2, i = 1, 2$. To follow the same development as above we first note that

$$E(s_i^2 - s_{12}) = (1-\rho)\sigma_e^2,$$

where ρ is the forecast error correlation coefficient. On expanding expressions for the variances and covariance of $s_i^2 - s_{12}, i = 1, 2,$ and collecting terms, we then obtain

$$\begin{aligned}\mathrm{var}(\hat{k}_o) &\approx \frac{1}{16(1-\rho)^2\sigma_e^4}[\mathrm{var}(s_1^2) + \mathrm{var}(s_2^2) - 2\,\mathrm{cov}(s_1^2, s_2^2)] \\ &= \frac{1}{(1-\rho)^2}\mathrm{var}(\hat{k}'). \end{aligned} \tag{A.2}$$

Or, more directly, we observe that the expression in square brackets in equation (A.1) is the variance of $s_1^2 - s_2^2$, and that this is equal to the variance of $(s_1^2 - s_{12}) - (s_2^2 - s_{12})$. As the errors of competing forecasts are in general positively correlated, the estimation variance of the optimal weight can be expected to exceed that of the inverse MSFE weight.

To implement expressions (A.1) and (A.2) for the forecasts used in our Monte Carlo study, we evaluate the terms in square brackets for the assumed Gaussian AR(2) data-generating process. The asymptotic variance of the sample variance of a normally distributed autocorrelated series with autocorrelation coefficients ρ_j is

$$\mathrm{var}(s^2) \approx \frac{2\sigma^4}{R}\sum_{-\infty}^{\infty}\rho_j^2.$$

This result is given in many time-series texts, whose authors usually cite Bartlett (1946). In this article, Bartlett stated a more general result, for the covariance of two sample autocovariances, and he subsequently gave an outline derivation in Bartlett (1955, §9.1). Following the same approach gives the covariance term we require as

$$\mathrm{cov}(s_1^2, s_2^2) \approx \frac{2}{R}\sum_{-\infty}^{\infty}\gamma_{12}^2(j),$$

where $\gamma_{12}(j)$ is the cross-lagged covariance function of e_1 and e_2. So, altogether we have

$$\mathrm{var}(\hat{k}') \approx \frac{1}{8R}(\mathrm{ACS}_1 + \mathrm{ACS}_2 - 2 \times \mathrm{CCS})$$

where ACS_i is the (doubly infinite) sum of squared autocorrelation coefficients of series e_{it}, $i = 1, 2$, and CCS is the corresponding sum of their squared cross-lagged correlation coefficients. In the set-up of our Monte Carlo experiments, these coefficients are obtained from the autocovariance-generating function of the AR(2) process for y_t and the related generating functions for the filtered series $e_{it} = h_i(L)y_t$. The infinite sums are truncated once the individual terms are sufficiently small, and possible sensitivity of the final result to the truncation point is checked.

The resulting 'theoretical' standard deviations of the distributions of the two estimators are shown in Table A.1 for a selection of the parameter values used in our experiments, alongside their 'empirical' counterparts calculated from the simulation sample distributions. With 1,000 independent replications at each parameter

combination, the standard error of the estimated standard deviation (SD) is equal to $\mathrm{SD}/\sqrt{2000} = 0.022 \times \mathrm{SD}$. With this in mind, the large-sample approximation is seen to provide reliable guidance to the simulation results for sample sizes as small as 30, for most of the parameter combinations considered. The approximation becomes less good as the autocorrelation of the forecast error series increases, and in these circumstances somewhat larger sample sizes are required before the equivalence is restored. Nevertheless, the theoretical analysis of this Appendix provides a more general support for a preference for $\hat{k}'$ over $\hat{k}_o$: neglecting the covariances of the competing forecast errors can be expected to lead to improved performance of combined forecasts based on estimated weights.

TABLE A1

Empirical and theoretical standard deviations of $\hat{k}'$ and $\hat{k}_o$

		$\hat{k}'$		$\hat{k}_o$	
ϕ_2	ρ	*Empirical*	*Theoretical*	*Empirical*	*Theoretical*
Case 1: $\phi_1 = 0.4$					
0.4	0.67	0.086	0.104	0.323	0.243
0.2	0.50	0.087	0.097	0.193	0.174
0.0	0.40	0.079	0.084	0.139	0.131
−0.2	0.33	0.069	0.070	0.107	0.102
−0.4	0.29	0.059	0.057	0.083	0.078
−0.6	0.25	0.048	0.044	0.065	0.058
−0.8	0.22	0.036	0.030	0.047	0.038
Case 1: $\phi_1 = 0.8$					
0.1	0.89	0.040	0.046	0.549	0.416
−0.1	0.73	0.053	0.057	0.228	0.208
−0.3	0.62	0.051	0.053	0.146	0.137
−0.5	0.53	0.046	0.045	0.103	0.096
−0.7	0.47	0.037	0.034	0.072	0.064
−0.9	0.42	0.027	0.019	0.046	0.032
Case 2: $\phi_1 = \phi_2$					
−0.9	0.57	0.043	0.037	0.106	0.086
−0.7	0.71	0.051	0.049	0.184	0.172
−0.5	0.83	0.046	0.046	0.285	0.274
−0.3	0.93	0.033	0.032	0.480	0.468
−0.1	0.99	0.012	0.012	1.374	1.347
0.1	0.99	0.014	0.014	1.242	1.219
0.3	0.87	0.044	0.044	0.351	0.343
Case 2: $\phi_1 = -\phi_2$					
0.3	0.87	0.042	0.044	0.347	0.343
0.1	0.99	0.013	0.014	1.205	1.219
−0.1	0.99	0.011	0.012	1.315	1.347
−0.3	0.93	0.030	0.032	0.456	0.468
−0.5	0.83	0.043	0.046	0.270	0.274
−0.7	0.71	0.048	0.049	0.175	0.172
−0.9	0.57	0.041	0.037	0.103	0.086