Forecast Combination when Outcomes are Difficult to Predict*

Graham Elliott
University of California, San Diego
9500 Gilman Drive
LA JOLLA, CA, 92093-0508

May 10, 2011

Abstract

We show that when outcomes are difficult to forecast in the sense that forecast errors have a large common component that (a) optimal weights are not affected by this common component, and may well be far from equal to each other but (b) the relative MSE loss from averaging over optimal combination is small. Hence researchers could well estimate combining weights that indicate that correlations could be exploited for better forecasts only to find that the difference in terms of loss is negligible. The results then provide another explanation for the commonly encountered practical situation of the averaging of forecasts being difficult to improve upon.

*Corresponding author : Graham Elliott, Department of Economics, 9500 Gilman Drive, LA JOLLA, CA, 92109, phone (858) 534 4481, fax (858) 534 7040, email:grelliott@ucsd.edu. Key words: Forecasting, Forecast Combination, Average Forecasts

1 Introduction

A basic puzzle in forecasting is the long documented good performance of simple averaging techniques for forecasts over combination methods that use correlations in the data to construct theoretically more precise forecast combinations. Even in their seminal paper introducing optimal forecast combination under mean square error (MSE) loss Bates and Granger (1969) noted this property. Indeed, it led them in this seminal paper to consider approaches that did not utilize the optimality theory completely, but instead consider less optimal methods. Clemen (1989), in a review of literature up to that time found that simple methods (such as averaging) often perform better than more complicated methods (such as the Bates and Granger optimal combination). In a more recent review Timmermann (2006) notes that in practice modifications to the combination methods (such as shrinkage) can outperform averaging (which both outperform optimal combination). However in a large scale forecasting study of US macroeconomic data Stock and Watson (2001) present results that show that averaging forecasts does as well as more complicated methods (they do not consider the Bates and Granger optimal combination, but another of the 'simpler' methods of that same paper).

The main thrust of the literature in explaining this puzzle has focussed on the effects of estimation error in estimating the weights for forecast combination (see Smith and Wallis (2009) for a recent example). Since simple averaging does not require any estimation, we expect that such estimation error harms the performance of methods where estimation of the weights is required relative to averaging. However the magnitudes of these effects are often not large. In a classical regression situation (stable parameters and variances), estimation error is of the order of $(m-1)/T$ when we have m forecasts to combine and T observations to estimate the weights. This is often relatively small. A number of authors (Bates and Granger (1969), Clemens and Hendry (2004)) have suggested that instability in the weights might also result in difficulties for situations where the weights are estimated. This will increase the error for such methods, although there is little documentation of actual rejections of tests of stability of the weights.

These possible explanations undoubtedly are part of the story. This paper examines another part of the explanation. We examine the potential gains from optimal combination over averaging the forecasts. If such gains are likely to be small for reasonable models, then (a) it is unlikely that more sophisticated methods can improve forecasts over simply averaging, and (b) it would provide an explanation of why even small amounts of estimation error might result in a better performance of averaging in practice. Situations that can be identified to have small gains would then be suspected to be more likely to be situations where we might just simply average and also could be used to examine if indeed these are the situations for which the puzzle arises.

We find that the larger the unpredictable component, the smaller is the potential gain from exploiting correlations in the data to provide more optimal forecast combination procedures over simple averages. An interesting piece of the result is that even though this is true there is no implication for the optimal combining weights — these weights can be far from the averaging weights even though the losses associated with the different weights are very similar. So in applied work a researcher might see estimated weights that give a very high weight to some forecasts and a low weight to others, suggesting that there are gains from exploiting the correlations, however in the context of MSE loss these gains are very small.

There are a number of practical implications for forecasters, as well as insights that can be gained from reevaluating the literature. First, at longer horizons variables are often harder to predict. In this sense we might expect to see that it is more difficult to find methods for forecast combination that use the correlations in the data that work better than averaging. On the other hand for very short term forecasting we might expect that such gains from utilizing the data in a more sophisticated manner might well yield useful gains. It will also be the case that for variables that are very hard to forecast, for example exchange rates, that there are few gains to optimal combination. Situations which are difficult to forecast will appear as ones where forecasts cluster (display herding) relative to the actual outcome, a situation that has been noted by many researchers. Ito (1990) details how in forecasts of the US Dollar Yen exchange rate market participants tend to be all on one side of the outcome (typically underpredicting changes in the exchange rate). The herding feature of forecasts of company earnings has been well documented in the finance literature (see for example Trueman (1994)).

The following section reviews quickly forecast averaging and optimal combination under MSE loss. The main results of the paper are then presented. Section 3 examines the results in a Monte Carlo. In section 4 we examine three forecasting methods of some prominent

US macroeconomic series. Using different forecast horizons as a 'natural experiment' for variation in the degree of predictability we are able to examine the theoretical results of the second section. We show that the effect does indeed show up in the data. A final section concludes.

2 Gains from optimal combination

We are interested in forecasting the variable y_{T+h}, and we decompose this into a potentially forecastable component $m_T = E[y_{T+h}|\Omega_T]$ and an unforecastable component ε_{T+h}, i.e.

$$y_{T+h} = m_T + \varepsilon_{T+h}. \tag{1}$$

The sigma field Ω_T is generated by a set of variables available to forecasters at time T, and as such is the best any forecaster in the set of forecasters can do in forecasting the outcome variable. Typical forecasters will have only a subset of this information — indeed we assume that each forecaster has differing variables they consider useful as well as potentially different models and estimation techniques. Because of this, forecasters attempting to minimize mean square loss provide forecasts of y_{T+h} (or equivalently estimates of m_T, since minimizing MSE is equivalent to attempting to find the conditional expectation for y_{T+h}) that differ amongst the forecasters[1]. Such differences are of course always found empirically. Further, following the assumption on Ω_T we assume that $E[\varepsilon_{T+h}|\Omega_T] = 0$ with variance σ^2_ε.

We have available a set of forecasts f_{iT} for $i = 1, ...m$. We assume that individual forecasters measure m_T with an idiosyncratic error. We write this as

$$E[y_{T+h}|\Omega_T] = f_{iT} + v_{iT+h} \tag{2}$$

where the timing convention on the error term is simply to keep track of the outcome for which the forecast is constructed. In the spirit of Bates and Granger (1969) as well as much of the forecast combination literature we assume that $E[v_{iT+h}] = 0$, which is to say that the forecasts are unbiased forecasts of the outcome. Let the variance covariance matrix of the idiosyncratic components be $\tilde{\Sigma}$.

[1] Alternatively, we could consider that forecasters do in fact observe the relevant information but have forecasts that differ from each other as they add in an additional random error. This would leave the following mathematical results unchanged, although is perhaps a less defendable story for the differences amongst forecasts.

From these two equations the forecast errors for each individual forecast have the form

$$y_{T+h} - f_{iT} = \varepsilon_{T+h} + v_{iT+h} \tag{3}$$

This is equivalent to the model of forecast errors of Zellner and Palm (1992). Define the vector of forecasts $f_T = (f_{1T}, ..., f_{mT})'$. We can write the variance covariance matrix of the forecast errors as

$$E[(y_{T+h}\iota - f_T)(y_{T+h}\iota - f_T)'] = \Sigma = \sigma_\varepsilon^2 \iota\iota' + \tilde{\Sigma} \tag{4}$$

where ι is an $mx1$ vector of ones.

For such a model, the 'forecastability' of y_{T+h} corresponds to the relative variances of m_T and ε_{T+h}. Holding the variance of m_T constant then it follows that the larger the variance of the unforecastable component σ_ε^2 the more difficult it is to forecast y_{T+h}. Throughout the remainder of this paper we hold the variance of the forecastable part constant and hence refer to forecastability and the size of σ_ε^2 interchangeably.

The first main result of the paper is that the optimal combination of forecasts, in the sense of Bates and Granger (1969), is independent of the degree of forecastability, i.e. independent of σ_ε^2. This result is stated in the following proposition and proved in the appendix.

Proposition 1 *For psd Σ write $\Sigma = \sigma_\varepsilon^2 \iota\iota' + \tilde{\Sigma}$, then for $\tilde{\Sigma}$ nonsingular then the Bates and Granger (1969) optimal weights from using Σ and $\tilde{\Sigma}$ are equivalent, i.e*

$$\begin{aligned} \omega^{opt} &= (\iota'\Sigma^{-1}\iota)^{-1}\Sigma^{-1}\iota \\ &= (\iota'\tilde{\Sigma}^{-1}\iota)^{-1}\tilde{\Sigma}^{-1}\iota. \end{aligned}$$

This result has a number of implications.

First, the proposition shows that the optimal Bates and Granger weights do not vary with the existence or size of a common component in the variance of the forecast errors σ_ε^2. This means that regardless of the level of predictability of the outcome, the Bates and Granger weights will still depend only on the idiosyncratic variability of forecast errors. Hence even with a very large variance of the unpredictable component, the weights may still well result in weights far from averaging if the forecasts are such that there are gains from a non even weighting in population.

Second, this result provides a potential explanation for situations where we observe high correlations between forecast errors but we might still have very non even weights on individual forecasts, even in population. The high correlation is driven by the unforecastable component being large, however the weights are still driven by the variations in the correlations. To see this notice that the correlations in Σ are equal to

$$\rho_{ij} = \frac{1 + (\tilde{\Sigma}_{ij}/\sigma_\varepsilon^2)}{\sqrt{\left(1 + (\tilde{\Sigma}_{ii}/\sigma_\varepsilon^2)\right)\left(1 + (\tilde{\Sigma}_{jj}/\sigma_\varepsilon^2)\right)}}$$

which converges to one as σ_ε^2 gets large (holding $\tilde{\Sigma}$ fixed).

Third, a model such as described in (1) and (2) can easily generate models for which forecast errors are very correlated however the optimal weights are equal to the combining weights from averaging. Such models, where the variances of the forecast errors are equal to each other ($\Sigma_{ii} = \Sigma_{jj}$ for all i, j) and the correlations are also equal to each other ($\Sigma_{ij} = \Sigma_{kl}$ for all $(i, j) \neq (k, l)$) are well known and often cited as a potential explanation for the good performance of using the average to combine forecasts. To see this consider a situation where the idiosyncratic errors are uncorrelated, all with the same variance, i.e. $\tilde{\Sigma} = \sigma_v^2$. Now we have a variance covariance matrix of the forecast errors that has covariance equal to $\sigma_\varepsilon^2/(\sigma_v^2 + \sigma_\varepsilon^2)$, so for $\sigma_\varepsilon^2/\sigma_v^2$ large enough the forecast errors could be highly correlated even though in the optimal combination all weights are equal to each other

Fourth, even when the variance covariance matrix of the forecast errors is very highly correlated, it may not mean that the reason for this is that the forecasts themselves are very similar. It could well be (and potentially mostly is) that the high correlation is due to the fact that none of the forecasts are picking up much of the variation in the variable to be forecast, but they all have a common source of noise. We often observe in practice that forecasts are clustered on one side of the outcome each period. One explanation of this is that much of the variation in y_{T+h} is actually due to unforecastable shocks (ε_T). The forecasts may well be evenly clustered around the conditional mean of y_{T+h} but the common shock ε_{T+h} causes them all to be off similarly. Thus when forecasts are compared to y_{T+h} they all look off in a similar way — the similarity is the unpredictable component.

Having established that the optimal weights are not affected by the variance of the unpredictable component, we now turn to the effect this variance has on the gains from optimal combination over simple averaging. The population loss from averaging over the forecasts when the forecast errors have a variance covariance matrix Σ is $m^{-2}(\iota'\Sigma\iota)$. For the Bates and Granger optimal combination, this loss is $(\iota'\Sigma^{-1}\iota)^{-1}$. Naturally, due to the latter being optimal, we have that $(\iota'\Sigma^{-1}\iota)^{-1} \leq m^{-2}(\iota'\Sigma\iota)$ for any positive definite variance covariance matrix Σ. The previous expression attains equality when the optimal weights are all equal to each other. It follows directly from these expressions as well as (4) that

$$m^{-2}(\iota'\Sigma\iota) = m^{-2}(\iota'[\sigma_\varepsilon^2\iota\iota' + \tilde{\Sigma}]\iota) = \sigma_\varepsilon^2 + m^{-2}(\iota'\tilde{\Sigma}\iota)$$

in the case of using average weights and

$$\begin{aligned}
(\iota'\Sigma^{-1}\iota)^{-1} &= (\iota'\left[\sigma_\varepsilon^2\iota\iota' + \tilde{\Sigma}\right]^{-1}\iota)^{-1} \\
&= \left(\iota'\tilde{\Sigma}^{-1}\iota - (1+\sigma_\varepsilon^2\iota'\tilde{\Sigma}^{-1}\iota)^{-1}\sigma_\varepsilon^2(\iota'\tilde{\Sigma}^{-1}\iota)^2\right)^{-1} \\
&= \left(\frac{\iota'\tilde{\Sigma}^{-1}\iota}{1+\sigma_\varepsilon^2\iota'\tilde{\Sigma}^{-1}\iota}\right)^{-1} \\
&= \sigma_\varepsilon^2 + \left(\iota'\tilde{\Sigma}^{-1}\iota\right)^{-1}
\end{aligned}$$

in the case of the optimal weights.

Thus clearly the size of the variance of the unforecastable component impacts the MSE loss from using either average weights or from using the optimal weights. In each case σ_ε^2 provides a wedge over and above the loss from loss based on forecasting the conditional mean (i.e. the expressions based on $\tilde{\Sigma}$ rather than Σ). The form is not surprising given the definitional independence between the forecastable and unforecastable components.

Since both loss from averaging and loss from the optimal weights are a function of the variance of the unforecastable component, the question remains as to how these are affected differently. This is an interesting question because it will only be useful to use optimal weights if the expected gain outweighs the estimation error that is incurred in estimating the $(m-1)$ weights. Defining the relative loss as the loss from averaging as a proportion of the loss from using the optimal weights, we have that the minimum relative loss is one (in population). Hence relative loss is

$$rl = \frac{\sigma_\varepsilon^2 + m^{-2}(\iota'\tilde{\Sigma}\iota)}{\sigma_\varepsilon^2 + \left(\iota'\tilde{\Sigma}^{-1}\iota\right)^{-1}} - 1. \tag{5}$$

Proposition 2 *For any positive definite $\tilde{\Sigma}$ relative loss given by (5) is either (a) equal to zero for all σ_ε^2, or (b) strictly decreasing in σ_ε^2.*

Proposition 2 shows that the relative loss from averaging over optimal weights is declining with increases in the variance of the unforecastable component, unless it was already zero.

The positive implication of this result is that the less forecastable the variable of interest the smaller is the expected gain from using optimal weights. Hence, taking into account that one incurs estimation error when using the optimal weights in practice whereas there is no estimation error in constructing the weights for the average forecast, we might expect that the gains from using optimal weights are less likely to be apparent the harder is the outcome variable to forecast.

The results give a potential explanation for the poor performance of estimated optimal weights in practice. If the outcome to be forecast is sufficiently difficult to predict then we would expect that relative loss is small enough that it becomes smaller than the estimation error that arises through the estimation of the optimal weights. Indeed, for any size of estimation error and any set of optimal weights there exists a level of unpredictability for which this is true. This explanation differs greatly from previous work that tends to suggest that the optimal weights are just close to the average weights. For this explanation we do not require that optimal weights be near the average weights — they could be very different, even with some of the weights being negative. But the effect of the size of the unpredictable component is that it reduces the gains from using these weights in practice.

Finally, we turn to considerations of reasonableness of having the common error large relative to idiosyncratic error. First, note that especially in macroeconomics in finance but true more generally it is often considered that good forecasts that capture a significant portion of the variation in the variable to be forecast are rare. Indeed, examples abound of forecast models that have difficulty outpredicting the simple use of the current level. Examples along this line include oil forecasting (for example Alquist and Kilian (2010)) and exchange rate forecasting. Breakdowns in the Phillips curve suggest it may be the case for inflation forecasting as well.

3 Monte Carlo Analysis

We choose the design of the Monte Carlo experiment to demonstrate both of the theoretical results above. First, that weights stay the same even when σ_ε^2 changes and second that relative loss is decreasing as this parameter changes. We also show a third result of interest — that even though the variance of the unforecastable component is sufficiently large as to make the relative loss from averaging zero, tests of the hypothesis that the optimal weights differ from equality (the average weights) can still be rejected with reasonable frequency (power of the test is nontrivial even when there are no gains over averaging).

Since results depend only on the forecast error (3) and not the individual y_t and f_{it}, we generate directly the forecast errors ε_t and v_{it}. We generate ε_t from a normal distribution with mean zero and variance σ_ε^2. Setting the number of forecasts $m = 3$ the idiosyncratic errors v_{it} are drawn from a normal distribution with mean zero and variance covariance matrix $\tilde{\Sigma}$, where

$$\tilde{\Sigma} = \begin{pmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 5 & 0.2 \\ 0.2 & 0.2 & 5 \end{pmatrix}.$$

This design is chosen so that the optimal weight vector $(0.75, 0.125, 0.125)$ is far from the equal weights vector of one third on each forecast. We draw 80 observations splitting this sample in to a first half of observations that are used to estimate the weights and then make a one step ahead forecast based on both the estimated and average weights. We then recursively update the data from observations 41-80 at each step repeating the estimation of the weights on all data up to that point and constructing ones step ahead forecasts. We then estimate the MSE of both combination methods by averaging over the squared one step ahead forecast errors. This is repeated 10000 times for various σ_ε^2, and the results reported in Table 1.

Table 1 shows in the first column the choice of σ_ε, the standard error of the common component of the forecast error. The second column shows the average relative loss over the Monte Carlo replications, computed as the average MSE over the replications for the averaging method divided by the same outcome for the optimal combination method minus one. The third and fourth columns report the average estimated weight parameters (averaged over all of the estimates for each sample and each Monte Carlo replication). The final column reports the average rejection frequency over the Monte Carlo replications of a full sample test that the optimal weights are equal to the average weights.

First, the effect presented in Proposition 1 that relative loss is decreasing in σ_ε^2 is shown in the second column. Larger variances of the unforecastable component result in the relative loss — which starts well above zero indicating a large loss from averaging — correspond with a smaller relative loss. For large enough variances the gain from optimal forecast combination is wiped out, and indeed the relative loss dips below zero giving an indication of the effects

Table 1: Monte Carlo Evaluatoin

σ_ε	rl	Estimated Weights		Rej Freq.
		ω_1	ω_2	
1	0.262	0.751	0.126	1.00
2	0.076	0.751	0.126	0.85
3	0.019	0.751	0.126	0.56
4	-0.004	0.751	0.127	0.36
5	-0.015	0.751	0.127	0.26
6	-0.021	0.751	0.128	0.21
7	-0.025	0.751	0.129	0.18

Notes: rl is average relative loss, Estimated weights are averages of out of sample estimated weights averaged across Monte Carlo simulations, Rej. Freq. is the proportion of MC replications where we reject the null hypothesis of equal weights at a 5% test using the full sample.

of estimation error on the optimal forecast combination estimation.

Second, the third and fourth columns make clear that the estimated weights are, as noted in the first proposition, unaffected by the size of the unforecastable component. The proposition makes clear that the population values are unaffected, and consistency of the parameter estimates ensures that this property flows through to the estimated quantities. In each column the parameter estimates are on average very close to their population values. So we can have an equivalence of optimal combination methods and averaging even when the optimal weights are far from the average weights.

Finally, we see that the equivalence of optimal combination and averaging is not a consequence of sampling error in the parameter estimates. Tests for equal weights still have non trivial power to reject equal weights even when the associated losses are basically equivalent. Hence rejection of even weights is not necessarily an indication that equal weights will perform poorly.

Variations in the design of $\tilde{\Sigma}$ will change the results in a quantitative way but not a qualitative way so long as the optimal weights related to $\tilde{\Sigma}$ are not equal to 1/3. If they are then clearly there is no difference in the models.

4 Demonstration with US Macro Data

The results of section 2 suggest that — without estimation error and when the variance covariance matrix of the heterogenous component of forecast errors stays the same — that in situations where the outcome is more difficult to forecast then we might expect that loss from averaging and use of the optimal forecasts would be closer to one than for variables that are easier to forecast (all else held constant). In this section we construct forecasts from three models for various macroeconomic data series at different horizons illustrate this point.

It is the comparison across different horizons that yields a situation where we expect the unobservable component of the forecast error to change. Consider the AR(1) model $y_t = \rho y_{t-1} + u_t$ where ρ is known and $u_t \sim (0, \tau^2)$ and serially independent. Then for a one step ahead forecast we have an expected MSE of σ^2, for a two step ahead forecast this is $\sigma^2(1+\rho^2)$ and for a four step ahead forecast $\sigma^2(1+\rho^2+\rho^4+\rho^6)$. For ρ different from zero, it is clear that as the forecast horizon increases the size of the unforecastable component increases. This merely corresponds to the greater uncertainty of forecasting further into the future. Given the results above, we would expect then that it would be less likely that optimally weighted combined forecasts would perform better than the average of the forecasts as the horizon increases. Of course this is one effect — it may well be that for data with little serial correlation that there is little effect here (say in the above example when $\rho = 0$), and it may also be the case that the covariance of the forecasts changes at different horizons (our theoretical results hold this fixed).

The macroeconomic series we examine are GDP (real gross domestic product, id GDPC96), Investment (real gross private domestic investment, id GPDIC96), Consumption (real personal consumption expenditures, id PCECC96), Government Spending (real government consumption expenditures and gross investment, id GCE96), Imports (real imports of goods and services, id IMPGSC96), unemployment (id LNS14000), and the GDP deflator (gross domestic product implicit price deflator index, id GDPDEF). For the inflation data the index is transformed into annualized inflation rates, for all other series except unemployment we take natural logarithms of the data. All data is quarterly, from the Federal Reserve Database (except unemployment, which is from the Bureau of Labor Studies).

We employ three methods for constructing forecasts of each of these series, chosen to be similar enough so that we obtain the common effect of forecast clustering but distinct enough that combination is not trivially an average of the forecasts (which happens if the forecasts are nearly identical). The methods are the random walk forecast (adjusted by a constant), an autoregressive model and a double exponential smoothing model[2]. In each case forecasts are constructed recursively with an initial portion of the data used solely for estimation, then all data up to each time period used to construct one quarter ahead, 2 quarter ahead and 4 quarter ahead forecasts. In each case forecasts are constructed for all but the first 25% of the data available for each series.

The average combined forecast from the three models is then constructed (without a constant adjustment). For the Bates Granger combined forecast, we again need to set aside some data for estimation of the weights. Hence our evaluation sample covers the second half of the sample, with the second quarter of the sample used to construct the first set of weights (estimated by restricted least squares, which is numerically identical to the formulas given above), then the weights are recursively updated through the sample at each time period. We then examine the average MSE's for each of the combining methods at each of the forecast horizons.

Figure 1 shows that the three forecasting methods provide one year ahead forecasts for GDP that 'herd', as is often seen in forecasts available for combination. When one forecast has an error below or above zero, for the most part they all do. It is precisely this effect that suggests a large σ_{ε}^2, as there is a common shock that all of the methods fail to pick up. The effects on forecasting of the great moderation (which may impact stability of the matrix Σ) are also clear.

[2]For all series except inflation we use a smoothing parameter of 0.96 for the levels and 0.2 for the differences, for inflation the latter parameter is set to zero.

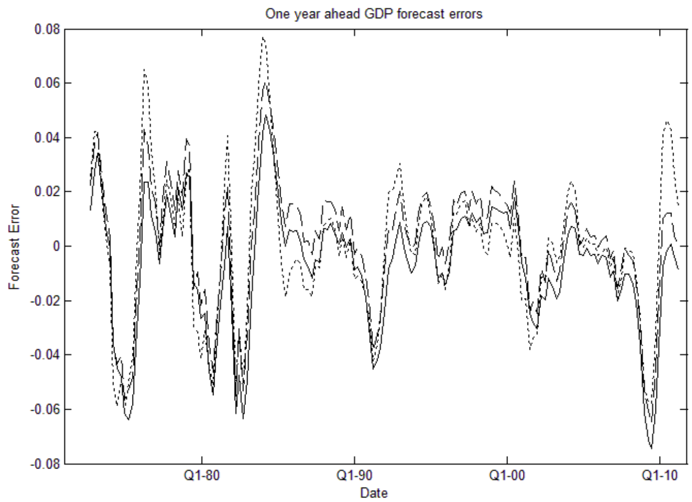

Table 2 presents the main results. We report the relative loss from averaging (5) where the population estimates of the MSE's are replaced by those estimated as in the previous paragraph. In brackets under each reported relative loss we report the t statistic testing the null that the squared forecast errors are equivalent (Diebold and Mariano (1995) DM test using MSE loss functions). These t statistics have an asymptotic normal distribution[3]. A statistically significant difference suggests that the deviation of relative loss from zero is not due to randomness. A positive value indicates that the average loss has a smaller MSE than the optimal weight combination loss, a negative value the reverse (so relative loss is positive when the statistic is negative). The last two columns report average out of sample MSE (over each set of out of sample observations and over all MC replications) at the one and two quarter horizons as a proportion of the same value at the four quarter horizon. A number less than one indicates that the MSE is smaller for the shorter horizons.

It is indeed the case in our sample that MSE loss is increasing in the forecast horizon, so it appears that the comparison across quarters is reasonable. For example for real GDP the average MSE for the one quarter ahead forecast is just 10% of the average MSE for the one year ahead forecast. The two quarter ahead forecast to one year ahead ratio is 30%. Results are qualitatively and for the most part quantitatively similar for all of the series apart from

[3] Whilst not formally justified here, it is expected that this approximation is likely to be reasonable. First, so long as the optimal weights are not the average weights the two combined forecast models will not have a degenerate spectral density at frequency zero (so issues of nested models do not arise). Second, we have used the same loss function (squared loss) in the forecast combination as the evaluation.

the inflation rate, where the one quarter ahead average MSE is 73% of the one year ahead average MSE (and for the two quarter ahead forecast this ratio is 85%).

Table 2: Relative Loss for Macroeconomic Series

	Relative Loss From Averaging			Forecastability	
	1qtr rl	2 qtr rl	4 qtr rl	MSE 1qtr	MSE 2 qtrs
GDP	$-0.002(0.02)$	$-0.045(1.49)$	$-0.127(1.70)$	0.10	0.30
Investment	$-0.039(1.29)$	$-0.083(1.30)$	$-0.101(0.85)$	0.13	0.36
Consumption	$-0.035(2.02)$	$-0.049(1.54)$	$-0.264(1.65)$	0.18	0.37
Government Spending	$0.003(-0.10)$	$-0.009(0.15)$	$-0.089(0.78)$	0.26	0.52
Imports	$-0.019(0.73)$	$-0.011(0.21)$	$0.009(-0.09)$	0.13	0.38
Unemployment	$0.141(-1.15)$	$0.122(-1.33)$	$0.026(-0.42)$	0.09	0.29
Inflation	$0.073(-1.51)$	$0.062(-1.23)$	$-0.041(0.323)$	0.73	0.85

Notes: Relative loss from averaging is the ratio of estimated average out of sample loss from averaging to optimal weights minus one. Forecastability measures are the average MSE from averaging at one and two quarter horizons relative to the four quarter horizon.

The results reinforce the strong performance of taking the simple average. In nearly all cases this is from a statistical perspective as good as optimal combination. This does not mean that in population there might not be in some cases some gain from optimal combination that is wiped out by sampling error in the estimation of the weights. However, despite this, the results of the paper do appear in a number of the series. In the majority of cases, we see that indeed the optimal combination is less useful relative to averaging in combining the forecasts at longer forecast horizons. The point estimates for GDP, investment, Government spending and unemployment indicate that the optimal combination forecast performs better at the shorter horizon than the longer horizon — a pattern expected from Proposition 2. The differences between the forecasts are not significantly different from one another by the DM test. At the longer horizons for each of these variables it appears that both optimal combination and averaging provide similarly useful forecast combination methods.

The anomalies are imports and inflation. For inflation, it seems that there is very little actual change in the size of the unforecastable component (the average MSE's across horizons increase but only slightly as the horizon increases). Hence this variable makes for a poor example of the feature discussed in this paper. For imports, it seems that at all horizons averaging and optimal combination yield essentially the same forecast. For the two and four quarter horizons the difference between the squared forecast errors is not only insignificant but the t statistics are close to zero.

Overall, the expected effect from the theoretical results appears in this data for most of the series in terms of the point estimates.

5 Conclusion

We show above that when there is a large unforecastable component of the outcome (or at least the available forecast models cannot capture a large common component of the outcome) then even though the optimal weights are not equal to the averaging weights, and can be sufficiently precisely estimated so that tests of equal weights would be rejected, it can still be the case that average weights result in MSE loss that is similar to the MSE loss that would arise through using optimal weights.

There are a number of implications that arise from the results. First, rejection of even weights does not in general imply that the optimal methods will do better than averaging the forecasts to provide a combination forecast. This is somewhat counterintuitive, so should help forecasters in thinking about what such pretests mean for their forecasting exercise. It also suggests that there are situations where methods such as pretesting for non equal weights to decide between optimal combination and averaging might not result in any gains at all. These would also be situations in which shrinkage methods are unlikely to provide any gain over simple averaging.

Second, in situations where it is clear that forecast models only capture a small amount of the variation in the outcome variable, it is likely that averaging the forecasts will yield close to all of the available gains in forecast combination. There are many forecasting situations such as this. Forecasting long horizons, or forecasting variables that are known to be highly variable such as exchange rates or oil prices, are examples of such situations.

Finally, from a theoretical perspective the results provide yet another situation in which averaging is likely to perform well. It is known that there exist many parameterizations of Σ for which averaging weights are indeed equivalent to optimal weights. Elliott (2010) shows that under restrictions on Σ (equal variances, nonnegative covariances and nonnegative weights) that the largest gains even when optimal weights are not even can be modest. The result here adds a realistic situation in which gains will be modest at best.

A number of authors have made the link between the forecast combination problem and mean-variance portfolio construction. So the results here have implications for this literature as well.

6 Appendix

Proof. (Proposition 1) We generalize slightly from the proposition, where above $b = 1$. From matrix inverse laws

$$\begin{aligned}\Sigma^{-1} &= \left(a\iota\iota' + b\tilde{\Sigma}\right)^{-1} \\ &= b^{-1}\left(\tilde{\Sigma}^{-1} - \frac{ab^{-1}}{1+ab^{-1}(\iota'\tilde{\Sigma}^{-1}\iota)}\tilde{\Sigma}^{-1}\iota\iota'\tilde{\Sigma}^{-1}\right).\end{aligned}$$

By direct calculation we have

$$\iota'\Sigma^{-1}\iota = b^{-1}(\iota'\tilde{\Sigma}^{-1}\iota)\left(\frac{1}{1+ab^{-1}(\iota'\tilde{\Sigma}^{-1}\iota)}\right)$$

and

$$\Sigma^{-1}\iota = b^{-1}(\tilde{\Sigma}^{-1}\iota)\left(\frac{1}{1+ab^{-1}(\iota'\tilde{\Sigma}^{-1}\iota)}\right)$$

so

$$\begin{aligned}\omega^{opt} &= (\iota'\Sigma^{-1}\iota)^{-1}\Sigma^{-1}\iota \\ &= \frac{b^{-1}(\tilde{\Sigma}^{-1}\iota)\left(\frac{1}{1+ab^{-1}(\iota'\tilde{\Sigma}^{-1}\iota)}\right)}{b^{-1}(\iota'\tilde{\Sigma}^{-1}\iota)\left(\frac{1}{1+ab^{-1}(\iota'\tilde{\Sigma}^{-1}\iota)}\right)} \\ &= (\iota'\tilde{\Sigma}^{-1}\iota)^{-1}\tilde{\Sigma}^{-1}\iota\end{aligned}$$

yielding the result. ■

Proof. (Proposition 2). When the optimal weights are equivalent to the averaging weights, i.e. $(\iota'\Sigma\iota)^{-1}\iota'\Sigma = m^{-1}\iota'$ then $rl = 0$ and is independent of σ_ε^2. If not we have by taking derivatives with respect to σ_ε^2 that

$$\begin{aligned}
\frac{\partial rl}{\partial \sigma_\varepsilon^2} &= \frac{1}{\sigma_\varepsilon^2 + \left(\iota'\tilde{\Sigma}^{-1}\iota\right)^{-1}} - \frac{\sigma_\varepsilon^2 + m^{-2}(\iota'\tilde{\Sigma}\iota)}{\left[\sigma_\varepsilon^2 + \left(\iota'\tilde{\Sigma}^{-1}\iota\right)^{-1}\right]^2} \\
&= \frac{\left(\iota'\tilde{\Sigma}^{-1}\iota\right)^{-1} - m^{-2}(\iota'\tilde{\Sigma}\iota)}{\left[\sigma_\varepsilon^2 + \left(\iota'\tilde{\Sigma}^{-1}\iota\right)^{-1}\right]^2} \\
&\leq 0
\end{aligned}$$

and the equality sign only holds in the case we are in the situation of (a). ■

References

ALQUIST, R., AND L. KILIAN (2010): "What do we learn from the price of crude oil futures," *Journal of Applied Econometrics*, 25, 539–573.

BATES, J., AND C. GRANGER (1969): "The Combination of Forecasts," *Operations Research Quarterly*, 20, 451–468.

CLEMEN, R. (1989): "Combining Forecasts: A Review and Annotated Bibliography," *International Journal of Forecasting*, 5, 559–581.

DIEBOLD, F., AND R. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–265.

ELLIOTT, G. (2010): "Averaging and the Optimal Combination of Forecasts," *unpublished mansucript*.

HENDRY, D., AND M. CLEMENTS (2004): "Pooling of Forecasts," *Econometrics Journal*, 1, 1–31.

ITO, T. (1990): "Foreign Exchange Rate Expectations: Micro Survey Data," *American Economic Review*, 80, 434–439.

PALM, F., AND A. ZELLNER (1992): "To Combine or not to Combine? Issues of Combining Forecasts," *Journal of Forecasting*, 11, 687–701.

Smith, J., and K. F. Wallis (2009): "A simple explanation of the forecast combination puzzle," *Oxford Bulletin of Economics and Statistics*, 71, 302–355.

Stock, J., and M. Watson (2001): "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," in *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive Granger*, ed. by R. Engle, and H. White, pp. 1–44. Oxford University Press, Oxford.

Timmermann, A. (2006): "Forecast Combinations," in *Handbook of Forecasting Volume 1*, ed. by G. E. et. al, pp. 135–196. Elsevier, Amsterdam.

Trueman, B. (2004): "Analyst forecasting and herding behavior," *Review of Financial Studies*, 7, 97–124.