

Linear Constraints and the Efficiency of Combined Forecasts

ROBERT T. CLEMEN

College of Business Administration, University of Oregon, U.S.A.

ABSTRACT

Studies of combined forecasts have typically constrained the combining weights to sum to one and have not included a constant term in the combination. In a recent paper, Granger and Ramanathan (1984) have argued in favour of an unrestricted linear combination, including a constant term. This paper shows that for the purpose of prediction it may make sense to impose restrictions on the combining model because of potential increases in forecasting efficiency. Empirical results show that small gains in forecasting efficiency can be obtained by restricting the linear combination of GNP forecasts from four econometric models.

KEY WORDS Combining forecasts Regression Linear constraints
Forecast efficiency and bias

In a seminal paper, Bates and Granger (1969) showed that a linear combination of forecasts can outperform the individual forecasts. Subsequent studies on combined forecasts have focused on simple averages of the individual forecasts (e.g. Makridakis and Winkler, 1983) or weighted averages where the weights are constrained to sum to one (Newbold and Granger, 1974; Nelson, 1972; Dickinson, 1975). Many published studies have shown that simple averages or averages that take into account the relative precisions of the forecasts can perform better than the individual forecasts. However, combination techniques that attempt to make use of sample information about the interdependence of forecasts have performed poorly in general, and many schemes have been proposed for improving the performance of such combinations. Granger and Newbold (1977, chap. 8) provide an overview.

In a recent article Granger and Ramanathan (1984) propose including a constant term and not restricting the weights to sum to one in the linear combination of forecasts. They state that the 'common practice of obtaining a weighted average of alternative forecasts should ... be abandoned in favour of an unrestricted linear combination *including a constant term*' (emphasis in original). Essentially, Granger and Ramanathan suggest that we perform a simple OLS regression with the actual value as the dependent variable and the forecast values as the independent variables.

The purposes of this paper are (1) to examine analytically the implications of Granger's and Ramanathan's proposal and (2) to subject their claim to an empirical test. The message of the

0277-6693/86/010031-08\$05.00

© 1986 by John Wiley & Sons, Ltd.

Received August 1985

Revised January 1986

paper is that it may indeed be appropriate to restrict the weights, or to require no constant term, or both, if the restricted combination results in a more efficient forecast. We show in the next section that there are potential gains in forecasting efficiency to be obtained by restricting the combinations, although the restricted combination may result in some bias in the forecast. Thus, the statistician must decide whether to accept some forecast bias for the sake of increased efficiency. In the second section, we present empirical results demonstrating that it may be worthwhile to impose the restrictions.

MODELS FOR COMBINING FORECASTS

For the following discussion we will assume that at time $t-1$ we have access to k forecasts, f_{1t}, \dots, f_{kt} , for θ_t , and that every forecast can be represented as a (possibly biased) estimate of θ_t ,

$$f_{it} = a_i + b_i \theta_t + u_{it} \quad (1)$$

where $u_{it} = (u_{1it}, \dots, u_{kit})^T$ has a normal distribution with mean vector $(0, \dots, 0)^T$ and covariance matrix Σ . Furthermore, we assume that the vectors u_{it} are independent and identically distributed from one time period to the next. Finally, we will assume that we have available past observations for times $t = 1, \dots, t-1$. Adopt the following notation:

$$[\theta, F] = \left[\begin{array}{c|ccc} \theta_1 & 1 & f_{11} & \dots & f_{k1} \\ \vdots & \vdots & \vdots & & \vdots \\ \theta_{t-1} & 1 & f_{1,t-1} & \dots & f_{k,t-1} \end{array} \right] \quad (2)$$

We include the vector of ones in the design matrix since in general we will be estimating regression coefficients including a constant term.

A regression model presupposes that we can write θ_t as a linear combination of the forecasts f_{1t}, \dots, f_{kt} :

$$\theta_t = \beta_0 + \beta_1 f_{1t} + \dots + \beta_k f_{kt} + \varepsilon_t \quad (3)$$

where the error term ε_t follows the ordinary least squares (OLS) assumptions. We can write this more compactly as

$$\theta_t = \mathbf{f}_t \boldsymbol{\beta} + \varepsilon_t \quad (4)$$

where $\mathbf{f}_t = (1, f_{1t}, \dots, f_{kt})$ and $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)^T$.

The OLS estimate of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \boldsymbol{\theta} \quad (5)$$

If we are then to forecast θ_t , given \mathbf{f}_t , let $\hat{\theta}_t$ denote the OLS predictor:

$$\hat{\theta}_t = \mathbf{f}_t \hat{\boldsymbol{\beta}} \quad (6)$$

We know that $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, and so $\hat{\theta}_t$ is an unbiased predictor of θ_t . Also, $\text{Var}(\hat{\theta}_t) = \sigma^2 \mathbf{f}_t (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{f}_t^T$, where $\sigma^2 = \text{Var}(\varepsilon_t)$. Furthermore, it is straightforward to show that

$$E(\theta_t - \hat{\theta}_t)^2 = \sigma^2 (\mathbf{f}_t (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{f}_t^T + 1) \quad (7)$$

Suppose, though, that we impose the restrictions that

$$\beta_0 = 0 \quad (8)$$

and

$$\beta_1 + \dots + \beta_k = 1 \quad (9)$$

Define

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

$$\boldsymbol{\gamma} = (0, 1)^T$$

We estimate $\boldsymbol{\beta}$ subject to the constraint that $\mathbf{R}\boldsymbol{\beta} = \boldsymbol{\gamma}$, using the same Lagrangian technique that Granger and Ramanathan employed, to find the constrained estimate $\boldsymbol{\beta}^*$:

$$\boldsymbol{\beta}^* = \hat{\boldsymbol{\beta}} - (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{R}^T [\mathbf{R}(\mathbf{F}^T \mathbf{F})^{-1} \mathbf{R}^T]^{-1} [\mathbf{R}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}] \quad (10)$$

Then the constrained prediction of θ_i is given by

$$\theta_i^* = \mathbf{f}_i \boldsymbol{\beta}^* \quad (11)$$

We see that

$$E(\theta_i^*) = \mathbf{f}_i \{ \boldsymbol{\beta} - (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{R}^T \mathbf{M} [\mathbf{R}\boldsymbol{\beta} - \boldsymbol{\gamma}] \} \quad (12)$$

and

$$\text{Var}(\theta_i^*) = \sigma^2 \{ \mathbf{f}_i [(\mathbf{F}^T \mathbf{F})^{-1} - (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{R}^T \mathbf{M} \mathbf{R} (\mathbf{F}^T \mathbf{F})^{-1}] \mathbf{f}_i^T \} \quad (13)$$

where $\mathbf{M} = [\mathbf{R}(\mathbf{F}^T \mathbf{F})^{-1} \mathbf{R}^T]^{-1}$. From this we can show that

$$E(\theta_i - \theta_i^*)^2 = \sigma^2 \{ \mathbf{f}_i [(\mathbf{F}^T \mathbf{F})^{-1} - (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{R}^T (\mathbf{I} - \sigma^{-2} \mathbf{M} [\mathbf{R}\boldsymbol{\beta} - \boldsymbol{\gamma}] [\mathbf{R}\boldsymbol{\beta} - \boldsymbol{\gamma}]^T) \mathbf{M} \mathbf{R} (\mathbf{F}^T \mathbf{F})^{-1}] \mathbf{f}_i^T + 1 \} \quad (14)$$

Comparing (14) and (7), we see that θ_i^* will have the smaller mean squared error as long as $\mathbf{L} = \mathbf{M} - \sigma^{-2} \mathbf{M} [\mathbf{R}\boldsymbol{\beta} - \boldsymbol{\gamma}] [\mathbf{R}\boldsymbol{\beta} - \boldsymbol{\gamma}]^T \mathbf{M}$ is a positive definite matrix. The first term of \mathbf{L} arises from the variance of $\boldsymbol{\beta}^*$ and the second term from the expectation of $\boldsymbol{\beta}^*$. Clearly, if $\mathbf{R}\boldsymbol{\beta} = \boldsymbol{\gamma}$, then the second term is a matrix of zeros. As a result $\mathbf{L} = \mathbf{M}$ and $E(\theta_i - \theta_i^*)^2 < E(\theta_i - \hat{\theta}_i)^2$ since \mathbf{M} is positive definite. However, if $\mathbf{R}\boldsymbol{\beta} \neq \boldsymbol{\gamma}$ (the imposed restrictions are inappropriate) it is possible that \mathbf{L} could be negative definite; the bias incurred by imposing the restrictions would more than offset the improved efficiency of θ_i^* , and thus the performance of θ_i^* would be worse (in terms of mean squared error) than $\hat{\theta}_i$.

A sufficient condition for $\mathbf{R}\boldsymbol{\beta} = \boldsymbol{\gamma}$ is that all k of the forecasts are unbiased, or $a_i = 0$ and $b_i = 1$ for $i = 1, \dots, k$. Although this assumption may often not be reasonable, if it is true then the regression can be written

$$\begin{aligned} \theta_i &= \beta_0 + \beta_1(\theta_i + u_{1i}) + \cdots + \beta_k(\theta_i + u_{ki}) + \varepsilon_i \\ &= \beta_0 + (\beta_1 + \cdots + \beta_k)\theta_i + (\beta_1 u_{1i} + \cdots + \beta_k u_{ki}) + \varepsilon_i \end{aligned}$$

For this to hold for every θ_i and (u_{1i}, \dots, u_{ki}) , we must have $\varepsilon_i = -(\beta_1 u_{1i} + \cdots + \beta_k u_{ki})$ and (8) and (9).

These results suggest that, if we believe that our forecasters are unbiased, we should constrain the linear combination of the forecasts. If we believe that they are 'almost' unbiased, we may still want to constrain the combination in order to gain increased efficiency even though this gain may be slightly offset by the incurred bias. In the context of combining forecasts we may want to impose (8) or (9) or both, depending on the characteristics of the underlying forecasts. In the next section we show that restricting the combination of U.S. GNP forecasts can improve the combined forecast performance compared to the unrestricted combination.

Wharton Econometrics (Wharton), Chase Econometrics (Chase), Data Resources, Inc. (DRI) and the Bureau of Economic Analysis (BEA) make quarterly forecasts of many economic variables. We used their level forecasts of GNP in real dollars (1970–1982) (obtained directly from Wharton and BEA and from the *Statistical Bulletin* published by The Conference Board for Chase and DRI) to construct growth rate forecasts (in percentage terms), and we calculated the deviations from actual growth as determined from GNP reported in *Business Conditions Digest*. Forecasts over four different horizons (1, 2, 3 and 4 quarters) were analysed. For example, the one-quarter forecast made in January predicts the percentage change (annual rate) of GNP from January to March, whereas the four-quarter forecast made in January predicts the percentage change that will occur from October to December of that year.

BEA makes only one forecast per quarter, usually early in the quarter, based in part on the first report of GNP for the previous quarter. Since Wharton, Chase and DRI update their forecasts each month, we attempted to assemble forecasts from them that were comparable to BEA's in timing and data used in making the forecast.

The data covered the 1970–1982 period. There were 46 observations available for each of the four forecast horizons. Each observation consisted of four forecasts and the actual value.

Exhibit 1 shows OLS estimates of the a_s and b_s for each forecaster and each of the four forecast horizons. These estimates were made using roughly the first half of the data. (Sample sizes vary due to missing data at the beginning of the series for Chase and DRI.) This allows an assessment of the forecasters before we begin evaluating the combined forecasts. Thus, we have acted as if we were at the beginning of 1976 (roughly) and were about to begin combining forecasts for the next period; a typical decision-maker in such a situation would (or at least should) have been interested in the recent performance of the forecasters.

The essential message conveyed by Exhibit 1 is that the a_s and b_s do not appear to be zeros and ones uniformly for all forecasters for any forecast horizon. Since the forecasters generally appear to be biased, both in terms of scale and a constant mean error, the combining method proposed by Granger and Ramanathan would appear to be appropriate.

Worth noting in Exhibit 1 is the overall decrease in R^2 as the forecast horizon increases, reflecting the notion that forecasting becomes less precise the further into the future one tries to forecast. Also, the Durbin–Watson statistic appears to decrease in general as the forecast horizon increases, which we may interpret as implying that serial correlation is more likely to be a problem for more distant predictions. This is an intuitive result. For example, two successive forecasts over long and overlapping periods would both be subject to any random shock that occurred after both forecasts were made and before the forecast random variables were realized.

Beginning with observation 24 and continuing through to the final one, the forecasts were combined on the basis of the preceding 20 observations. For example, in evaluating Granger's and Ramanathan's proposal we used observations $t-21$ to $t-1$ as the design matrix in a simple OLS regression to estimate the coefficients for combining forecasts made. Thus, the combined forecasts are 'pure' forecasts in the sense of Klein (1984); they represent combined forecasts that might have been made and used by decision makers at time t .

Exhibit 2 shows mean squared errors (MSE) and mean absolute errors (MAE) for four combination methods. Let constraint 1 indicate the constraint that $\hat{\beta}_0 = 0$ (no constant) and constraint 2 that $\hat{\beta}_1 + \dots + \hat{\beta}_4 = 1$. Then the four methods are (I) OLS (unconstrained); (II) OLS with constraint 1; (III) OLS with constraint 2; (IV) OLS with both constraints.

Even though Exhibit 1 would lead us to believe that method I is the 'correct' model, it is evident from Exhibit 2 that there are gains to be made by including constraints in the combining

	Wharton	Chase	DRI	BEA
1-quarter horizon				
$a_i(S_{ai})$	0.85(0.57)	0.86(0.51)	1.60(0.66)	1.35(0.49)
$b_i(S_{bi})$	0.71(0.10)	0.75(0.09)	0.64(0.11)	0.73(0.09)
R^2	0.70	0.81	0.65	0.76
d	1.71	2.48	1.99	1.89
n	23	20	20	23
2-quarter horizon				
$a_i(S_{ai})$	2.67(0.70)	2.53(0.66)	3.44(0.54)	2.49(0.55)
$b_i(S_{bi})$	0.44(0.12)	0.45(0.11)	0.37(0.09)	0.46(0.09)
R^2	0.38	0.48	0.49	0.51
d	1.90	1.78	1.84	1.90
n	23	20	20	23
3-quarter horizon				
$a_i(S_{ai})$	2.72(0.53)	3.57(0.47)	3.75(0.50)	3.47(0.52)
$b_i(S_{bi})$	0.29(0.09)	0.25(0.08)	0.20(0.09)	0.22(0.09)
R^2	0.35	0.35	0.23	0.24
d	2.38	1.67	1.67	2.22
n	23	20	20	23
4-quarter horizon				
$a_i(S_{ai})$	3.38(0.43)	4.08(0.48)	4.13(0.44)	4.38(0.45)
$b_i(S_{bi})$	0.17(0.07)	0.12(0.08)	0.14(0.07)	0.05(0.07)
R^2	0.21	0.11	0.16	0.02
d	1.30	1.20	1.45	1.17
n	23	20	20	23

Exhibit 1. Evaluation of individual forecasters

(a) Mean absolute errors (MAE)

Forecast horizon (quarters)	Wharton	Chase	DRI	BEA	I	II	III	IV	n
1	2.863	2.954	3.122	2.907	3.523	3.260	3.324	3.043	26
2	3.447	3.888	3.524	3.621	3.380	3.344	3.550	3.653	25
3	3.589	4.166	3.605	3.698	3.867	3.860	3.817	4.132	24
4	3.529	3.958	3.631	4.065	4.408	4.384	4.812	4.592	23

(b) Mean squared errors (MSE)

Forecast horizon (quarters)	Wharton	Chase	DRI	BEA	I	II	III	IV	n
1	11.630	12.529	14.291	12.359	18.733	14.974	14.977	13.024	26
2	17.919	21.985	17.120	18.883	21.377	18.299	21.034	19.607	25
3	22.012	25.991	20.091	24.495	26.539	25.059	25.955	27.035	24
4	23.593	26.432	20.939	25.815	28.626	27.849	33.384	31.517	23

- Method: (I) OLS
 (II) OLS with no constant
 (III) OLS with $\hat{\beta}_1 + \dots + \hat{\beta}_4 = 1$
 (IV) OLS with no constant and $\hat{\beta}_1 + \dots + \hat{\beta}_4 = 1$

Exhibit 2

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(a) MAE

x	y				row sum
	II	III	IV	I	
II		3	3	4	10
III	1		2	2	5
IV	1	2		1	4
I	0	2	3		5

(b) MSE

x	y				row sum
	II	IV	III	I	
II		3	4	4	11
IV	1		3	2	6
III	0	1		3	4
I	0	2	1		3

Exhibit 3. Number of times x (row) outperformed y (column). For example, from (a) we see that II outperformed IV three times in terms of MAE. The row sum (S) is an overall measure of how many pairwise comparisons the x method 'won' ($0 \leq S \leq 12$)

procedure. Exhibit 3 summarizes the performance results by indicating how often any given method outperformed the other methods. In terms of MAE and MSE both, it is clear that method II performed the best overall. Methods I, III and IV follow, although there is no clear ranking among the three. In terms of MSE, method I was outperformed more often than the other three methods.

In the preceding section we argued that imposing inappropriate constraints causes the improperly constrained combination to be biased. Exhibit 4 compares the mean errors among the four combination schemes. As predicted, in general the constrained methods yield mean errors with magnitude larger than method I. Only method III in horizons 3 and 4 and method II in horizon 4 had mean errors closer to zero than the corresponding mean error for method I. Also it is worth noting that a small mean error does not necessarily indicate small MAE or MSE; only in horizon 4 did the method with the smallest MAE and MSE also have the ME with the smallest magnitude.

Forecast horizon (quarters)	Combining method				n
	I	II	III	IV	
1	-0.838	-1.244	-1.294	-1.183	26
2	-0.004	-0.182	-0.316	0.100	25
3	-0.324	0.585	0.270	1.090	24
4	-1.012	-0.176	-0.470	1.077	23

Exhibit 4. Mean errors (ME)

The proposal has been made to abandon constrained linear combinations of forecasts in favour of unconstrained combinations on the grounds that the latter will yield better results. In making their proposal, Granger and Ramanathan focused on the smaller sum of squared errors that could be attained by using an unconstrained model. It is indeed true that unconstrained least squares will (by definition) result in the smallest sum of squared errors for the data used to fit the regression. The issue, however, is not to minimize the squared errors within the fitting data, but rather to improve the performance of the combination in forecasting outside the fitting data. We have shown how inclusion of constraints can improve forecast efficiency, and have suggested that in some cases the increased efficiency may more than offset any incurred bias.¹

The central notion presented here, that an increase in efficiency may be worth some added bias, is certainly not new. For example, ridge regression (Hoerl and Kennard, 1970) is often used to improve the efficiency of estimated regression parameters when multicollinearity is a problem. The efficiency comes at the price of some bias, but in multicollinearity problems this is a welcome trade-off. The results we have presented indicate that our combined forecasts also suffer from inefficiency in general. Exhibit 2 points this out dramatically. In general the combined forecasts did not perform substantially better than any of the individual forecasts, and in many cases the combinations perform worse than all of the individual forecasts. The imposition of constraints appears to have had some value in ameliorating this situation. The inefficiency of the combinations is a result of the multicollinearity in the individual forecast errors. For a detailed examination of this problem and a Bayesian model that improves the performance of a combined forecast, see Clemen and Winkler (1986).

Granger and Ramanathan state that the results in their example 'certainly indicate the superiority of Method C [OLS], although the amount of improvement is not always very large'. On the basis of our experiment, we would argue the opposite, but still with the same caveat. Perhaps more to the point is that both empirical tests involve small samples, a chronic problem in studies of combined forecasts. Simulation studies can overcome this problem and might provide valuable insight. For the purpose of making recommendations to practitioners, however, it seems more appropriate to focus on combining forecasts in as many real world situations as possible.

ACKNOWLEDGEMENTS

Data from the Wharton and BEA models were graciously provided by Donald Straszheim and George Jaszi, respectively. Bob Winkler, Sergio Koreisha and anonymous referees provided helpful comments on this topic.

REFERENCES

- Bates, J. M. and Granger, C. W. J., 'The combination of forecasts', *Operations Research Quarterly*, 20 (1969), 451-468.
- Clemen, R. T. and Winkler, R. L., 'Combining economic forecasts', *Journal of Business and Economic Statistics*, 4 (1986), 39-46.

¹ The problems here are not unique to Granger's and Ramanathan's study. Poor forecasting performance by OLS, constrained or not, outside the fitting data can also be caused by structural changes. In this case, incorporating such changes into the model might produce better combined forecasts.

- Dickinson, J. P., 'Some statistical results in the combination of forecasts', *Operations Research Quarterly*, **24** (1975), 253-260.
- Granger, C. W. J. and Ramanathan, R., 'Improved methods of combining forecasts', *Journal of Forecasting*, **3** (1984), 197-204.
- Granger, C. W. J. and Newbold, P., *Forecasting Economic Time Series*, New York: Academic Press, 1977.
- Hoerl, A. E. and Kennard, R. W., 'Ridge regression: biased estimation of nonorthogonal problems', *Technometrics*, **12** (1970), 55-67.
- Klein, L. R., 'The importance of the forecast', *Journal of Forecasting*, **3** (1984), 1-9.
- Makridakis, S. and Winkler, R. L., 'Averages of forecasts: some empirical results', *Management Science*, **29** (1983), 987-996.
- Nelson, C. R., 'The predictive performance of the F.R.B.-M.I.T.-PENN model of the U.S. economy', *American Economic Review*, **62** (1972), 902-917.
- Newbold, P. and Granger, C. W. J., 'Experience with forecasting univariate time series and the combination of forecasts', *Journal of the Royal Statistical Society, Series A*, **137** (1974), 141-164.

Author's biography:

Robert T. Clemen is an Assistant Professor of Business at the University of Oregon. He has a Ph.D. in Quantitative Business Analysis from Indiana University. His research interests are in combining forecasts, decision theory and Bayesian statistics, and he has published articles in *Operations Research*, *Management Science*, *Journal of Business Economics and Statistics*, and *Journal of Forecasting*.

Author's address:

Robert T. Clemen, College of Business Administration, University of Oregon, Eugene, OR 97403-1208, U.S.A.