25th International Congress on Modelling and Simulation, Darwin, NT, Australia, 9 to 14 July 2023
mssanz.org.au/modsim2023

Selecting the relevant forecasts for forecast combination

F. Chan and **R. Singh**

School of Accounting, Economics and Finance, Curtin University, GPO BOX U1987, Perth, Western Australia
Email: ranjodh.singh@curtin.edu.au

Abstract: The importance of producing accurate forecasts is vital for informed decision-making. One method of generating accurate forecasts is to combine multiple forecasts (if available) using forecast combination methods. Forecast combination methods make use of all available forecasts to generate a combined forecast. This includes individual forecasts that are poor performers. Given this, it may be sensible to exclude these poor performing forecasts i.e. only select relevant forecasts prior to combining them. In order to achieve this, a well-known shrinkage estimator is used to select the relevant forecasts which are then used to generate a forecast combination. This paper assesses if the forecast combinations resulting from the selected forecasts have better accuracy than combinations that include all forecasts. The results are dependent on the forecast combination method used. For simple forecast combination methods such simple average or the Bates and Granger method, the selection does improve the resulting forecast combination. However, for regression based forecast combination methods, selection does not make any significant difference.

Keywords: *Forecast combination, shrinkage, forecast performance*

1 INTRODUCTION

The importance of producing accurate forecasts is vital for informed decision-making. Accurate forecasts allows individuals to plan effectively, allocate resources efficiently and reduce uncertainty. Furthermore, if multiple forecasts are available, it is possible to combine these using forecast combination methods. Given that, each individual forecast captures some aspect of the true underlying process, combining them may improve forecast prediction. Indeed, there is evidence that combined forecasts tend to perform better that any single forecast on average (Chan & Pauwels (2018)). A key feature of combining individual forecasts is the weight assigned to each forecast. As such, forecast combination methods can be classified into two main categories (i) simple combination methods and (ii) regression based combination methods. Examples of simple combination methods include Simple Average (SA) (equal weights) and Bates and Granger (BG) (Bates & Granger (1969)). Regression based combination methods include ordinary linear squares (OLS) regression (Granger & Ramanathan (1984)) and the least absolute deviation (LAD) regression (Nowotarski et al. (2014)). For both regression based cases, the estimated coefficients are used as weights for individual forecasts. In practise, these methods take into account all of the individual forecasts in order to generate a forecast combination. However, some of these forecasts may be poor performers and as such it would be sensible to exclude them from the forecast combination. Hence, the selection of relevant individual forecasts is required prior to producing the forecast combination.

One possible method of selecting forecasts is to use a shrinkage estimator such as the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani (1996)). This estimator is used in the estimation of regression models where it reduces the impact of irrelevant variables and hence improves model interpretability (Chan & Mátyás (2022), Chan et al. (2022). This is achieved by shrinking the coefficients of less important variables towards zero and in some cases exactly zero. Thus, removing the variable from the model. Furthermore, there are other practical reasons to use LASSO in the forecasting context. Most macroeconomic forecasting data sets contain multiple individual forecasts and these forecasts (k) are greater than the number of observations (n) available i.e. $n << k$. In such cases, regression based forecast combination methods such as OLS and LAD cannot be used without selecting the individual forecasts first.

Based on the above, this paper aims to apply a shrinkage estimator to select relevant individual forecasts prior to combining them. In doing so, the paper examines if the accuracy of the resulting forecast combination (with selection) is better than the forecast combination produced without selection. This paper also compares the results based on LASSO to an alternative forecast combination method which is capable of selecting individual forecasts prior to combining them. This method is known as the Constrained Least Squares (CLS) regression. Unlike the OLS and LAD methods, this method constraints the weights to be greater than or equal to zero and also sum up to one (affine combination). Given this context, currently there are no known studies which examine the effect of selection on forecast accuracy across different forecast combination methods. As such, this paper will contribute to this gap in the current literature.

The structure of the paper is as follows; Section 2 describes the details on the data and methodology used to address the research question. Section 3 contains the interpretation and discussion of the results and lastly section 4 states the conclusions of the paper.

2 DATA AND METHODOLOGY

In order to address the research question, a Monte-Carlo experiment was conducted. The experiment consisted of generating a true data generating process/true model along with a number of individual forecasting models, each of which attempted to predict the true model. Each of the individual forecasting models were systematically lacking one or more components of the true model. The forecasting models were designed to reflect the stylised facts present in data sets containing multiple individual forecasts. These include (i) individual forecasts being correlated to the true data values (ii) individual forecasts being highly correlated with each other. To test the validity of the selection process, pure noise models were also added to the list of forecasting models. A description of all the models used in experiment is provided in Table 1.

As shown in Table 1, y_{0t} denotes the true model which is an auto-regressive moving average model with a single lag (ARMA(1,1)). All of other models (y_{1t} to y_{13t}) were used to generate forecasts for y_{0t}. The unconditional mean of all models was set to zero (a_0 = 0). The first three models (y_{1t} to y_{3t}) use previous/historical values of the true model to generate forecasts. The next three models (y_{4t} to y_{6t}) denote auto-regressive (AR) models with lags of 1, 2 and 3. Similarly, models for y_{7t} to y_{9t} denote moving average (MA) models with lags of 1, 2, and 3. The models for y_{10t} and y_{11t} denote ARMA(2,2) and ARMA(3,3) respectively and lastly

y_{0t}	$\alpha_0 + \alpha_1 y_{0t-1} + \beta_1 \varepsilon_{0t-1} + \varepsilon_{0t}$
y_{1t}	$\alpha_0 + \alpha_1 y_{0t-1}$
y_{2t}	$\alpha_0 + \alpha_1 y_{0t-1} + \alpha_2 y_{0t-2}$
y_{3t}	$\alpha_0 + \alpha_1 y_{0t-1} + \alpha_2 y_{0t-2} + \alpha_3 y_{0t-3}$
y_{4t}	$\alpha_0 + \alpha_1 y_{4t-1} + \varepsilon_{4t}$
y_{5t}	$\alpha_0 + \alpha_1 y_{5t-1} + \alpha_2 y_{5t-2} + \varepsilon_{5t}$
y_{6t}	$\alpha_0 + \alpha_1 y_{6t-1} + \alpha_2 y_{6t-2} + \alpha_3 y_{6t-3} + \varepsilon_{6t}$
y_{7t}	$\alpha_0 + \beta_1 \varepsilon_{7t-1} + \varepsilon_{7t}$
y_{8t}	$\alpha_0 + \beta_1 \varepsilon_{8t-1} + \beta_2 \varepsilon_{8t-2} + \varepsilon_{8t}$
y_{9t}	$\alpha_0 + \beta_1 \varepsilon_{9t-1} + \beta_2 \varepsilon_{9t-2} + \beta_3 \varepsilon_{9t-3} + \varepsilon_{9t}$
y_{10t}	$\alpha_0 + \alpha_1 y_{10t-1} + \alpha_2 y_{10t-2} + \beta_1 \varepsilon_{10t-1} + \beta_2 \varepsilon_{10t-2} + \varepsilon_{10t}$
y_{11t}	$\alpha_0 + \alpha_1 y_{11t-1} + \alpha_2 y_{11t-2} + \alpha_3 y_{11t-3} + \beta_1 \varepsilon_{11t-1} + \beta_2 \varepsilon_{11t-2} + \beta_3 \varepsilon_{11t-3} + \varepsilon_{11t}$
y_{12t}	$\alpha_0 + \varepsilon_{12t}$
y_{13t}	$\alpha_0 + \varepsilon_{13t}$

Table 1. Specification - Forecasting models

y_{12t} and y_{13t} are pure noise models with different variances. The error terms in the models are correlated to both the true model as well as the other individual forecasting models. Lastly, it is assumed that both the correlations and parameters across all models are fixed.

The first stage of the analysis consists of the following steps:

1. Generate $n = 1,000$ values for the true model
2. Generate forecasts from each of the 13 individual forecasting models
3. Produce SA forecast combinations using training data (80/20 split)
4. Validate the SA forecast combinations on the test data using the following accuracy metrics - Mean Error (ME), Root Square Mean Error (RMSE) and Mean Absolute Error (MAE)
5. Repeat the above steps 100, 250 and 500 times in order to assess the variability across the accuracy metrics.
6. Repeat the above steps across the remaining forecast combination methods - BG, OLS, LAD and CLS.
7. Repeat the entire analysis for a larger sample ($n = 10,000$)

The second part of the analysis consists of selecting individual forecasting models using LASSO. Only the selected models are then allowed to enter the forecast combination. As such, a two stage process was introduced - selection stage and combination stage. These following steps describe the two stage approach:

1. Use LASSO to select forecasting models using the training data
2. Use the selected subset of forecasting models to produce SA forecast combinations.
3. Validate the SA forecast combinations on the test data using the following accuracy metrics - Mean Error (ME), Root Square Mean Error (RMSE) and Mean Absolute Error (MAE)
4. Repeat the above steps for 100, 250 and 500 times in order to assess the variability across the accuracy metrics.
5. Repeat the above steps across the remaining forecast combination methods - BG, OLS, LAD and CLS.
6. Repeat the entire analysis for a larger sample ($n = 10,000$)

For both sets of analysis (with and without selection), the average of the performance metrics (ME, RMSE and MAE) are computed for each replication sample (100, 250 and 500) across both the smaller and larger data sets (1,000 and 10,000). Tables 2 and 3 contain the average performance metrics from all the MC experiments. In addition to this, the proportion of times a forecasting model was selected (using LASSO and CLS) was also recorded. Figures 1 and 2 illustrate the results for all replication samples across both small and large data sets. The entire analysis was conducted using R software R Core Team (2021) and the *ForecastComb* package Weiss et al. (2018).

3 RESULTS

As mentioned in the previous sections, the objective of this paper to the examine if selecting forecasts prior to combining them improves the accuracy of the forecast combination. Based on the results presented in Table 2, selection improved the accuracy for simple forecast combinations relative to regression based forecast combinations. The RMSE and MAE decreased for both SA and BG based combinations. However, for combinations based on OLS and LAD, both RMSE and MAE remain relatively unchanged. The results were fairly constant across all replication samples (100, 250 and 500). Finally, these results were also true for the large sample ($n = 10,000$ Table 3).

Tables 4 and 5 show the results for the CLS method. As mentioned previously, the CLS method is capable of selecting forecasts prior to combining them and as such can be viewed as an alternative to using LASSO based approach. The accuracy of the CLS combinations is better than both SA and BG based combinations with and without selection. However, the CLS based combinations do not outperform the forecasts combinations based on OLS and LAD. This is expected since the additional constraints will produce sub-optimal results for CLS relative to OLS. However, the CLS weights are interpretable compared to both OLS and LAD.

In addition to the above, LASSO was applied CLS outputs to investigate if the forecast accuracy could be further improved. This resulted in a *double selection* of forecasting models or in other words, 'CLS with selection'. The results of these exercise are presented in Tables 4 and 5. Based on this, there were no substantial changes to the average values of the accuracy metrics.

Lastly, Figures 1 and 2 show the results on selection consistency i.e. how often a forecasting model was selected for both LASSO and CLS across small and large samples. For both samples, both LASSO and CLS select models $1, 2, 4, 6, 7, 8, 10$ and 11 with similar proportions. However, there differences in selection frequency for models $3, 5, 9, 12$ and 13. In fact, the CLS method never selects model 3 (model which uses 3 lagged historical values). This could be due to the fact that the CLS method can account for higher correlation across individual forecasts compared to LASSO. This difference increases even more for the large sample. Model 5 is an interesting case. It is selected with increasing frequency by the CLS method relative to LASSO for the small sample. However, for the large sample, LASSO selects model 5 with increasing frequency relative to CLS. For model 9 selection differences do exist for the small sample, but these become negligible for the large sample.

With regard to pure noise models, the CLS method selects the models 12 and 13 with a higher frequency compared to LASSO for the small sample. However, this result changes for the large sample where both the CLS and LASSO select model 12 and 13 with very low frequencies.

	Reps = 100		Reps = 250		Reps = 500	
Accuracy	SA	SA with selection	SA	SA with selection	SA	SA with selection
ME	0.000	-0.001	-0.005	-0.004	-0.002	-0.003
RMSE	0.708	0.655	0.715	0.660	0.714	0.659
MAE	0.565	0.524	0.572	0.527	0.571	0.526
Accuracy	BG	BG with selection	BG	BG with selection	BG	BG with selection
ME	-0.001	-0.001	-0.004	-0.004	-0.003	-0.003
RMSE	0.655	0.646	0.662	0.651	0.661	0.648
MAE	0.523	0.516	0.530	0.520	0.528	0.518
Accuracy	OLS	OLS with selection	OLS	OLS with selection	OLS	OLS with selection
ME	-0.008	-0.008	0.000	0.000	-0.001	-0.001
RMSE	0.600	0.601	0.605	0.608	0.603	0.605
MAE	0.479	0.481	0.483	0.486	0.482	0.483
Accuracy	LAD	LAD with selection	LAD	LAD with selection	LAD	LAD with selection
ME	-0.007	-0.007	0.000	0.000	-0.001	-0.001
RMSE	0.600	0.600	0.605	0.608	0.603	0.605
MAE	0.479	0.481	0.483	0.485	0.482	0.483

Table 2. Small sample - Forecast accuracy and model selection

	Reps = 100		Reps = 250		Reps = 500	
Accuracy	SA	SA with selection	SA	SA with selection	SA	SA with selection
ME	-0.002	-0.002	0.002	0.001	0.002	0.002
RMSE	0.715	0.662	0.716	0.663	0.716	0.663
MAE	0.570	0.528	0.572	0.529	0.571	0.529
Accuracy	BG	BG with selection	BG	BG with selection	BG	BG with selection
ME	-0.002	-0.002	0.001	0.001	0.002	0.002
RMSE	0.662	0.653	0.663	0.654	0.662	0.654
MAE	0.528	0.521	0.529	0.522	0.528	0.522
Accuracy	OLS	OLS with selection	OLS	OLS with selection	OLS	OLS with selection
ME	-0.001	-0.001	0.001	0.001	0.001	0.001
RMSE	0.600	0.600	0.600	0.600	0.600	0.600
MAE	0.479	0.479	0.479	0.479	0.478	0.479
Accuracy	LAD	LAD with selection	LAD	LAD with selection	LAD	LAD with selection
ME	-0.001	-0.001	0.001	0.001	0.001	0.001
RMSE	0.600	0.600	0.600	0.600	0.600	0.600
MAE	0.479	0.479	0.479	0.479	0.478	0.479

Table 3. Large sample - Forecast accuracy and model selection

	Reps = 100		Reps = 250		Reps = 500	
Accuracy	CLS	CLS with selection	CLS	CLS with selection	CLS	CLS with selection
ME	-0.003	-0.003	-0.002	-0.002	-0.002	-0.002
RMSE	0.625	0.624	0.628	0.628	0.626	0.626
MAE	0.500	0.499	0.502	0.502	0.501	0.500

Table 4. CLS performance - Small sample

	Reps = 100		Reps = 250		Reps = 500	
Accuracy	CLS	CLS with selection	CLS	CLS with selection	CLS	CLS with selection
ME	-0.002	-0.002	0.001	0.001	0.002	0.002
RMSE	0.626	0.626	0.626	0.626	0.626	0.626
MAE	0.499	0.499	0.500	0.500	0.499	0.499

Table 5. CLS performance - Large sample

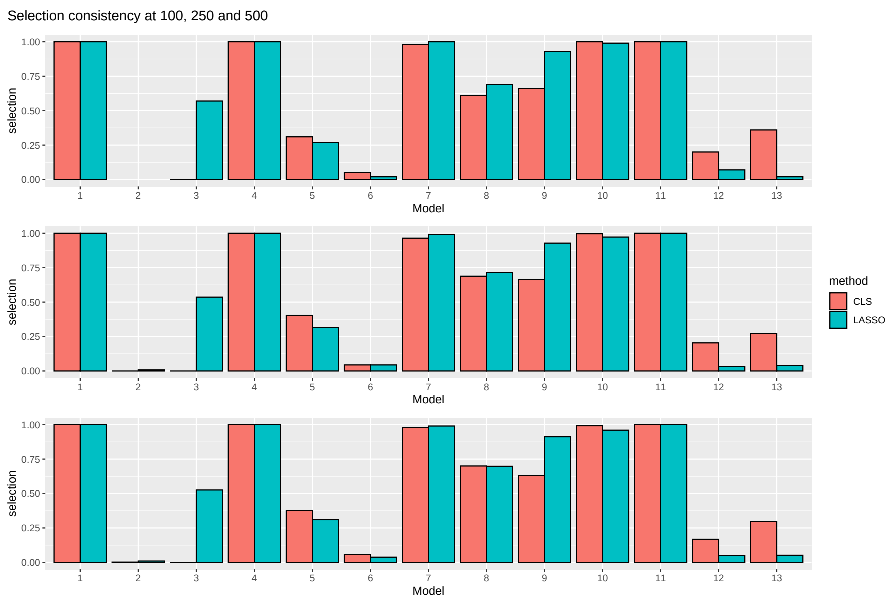

Figure 1. Model selection - small sample

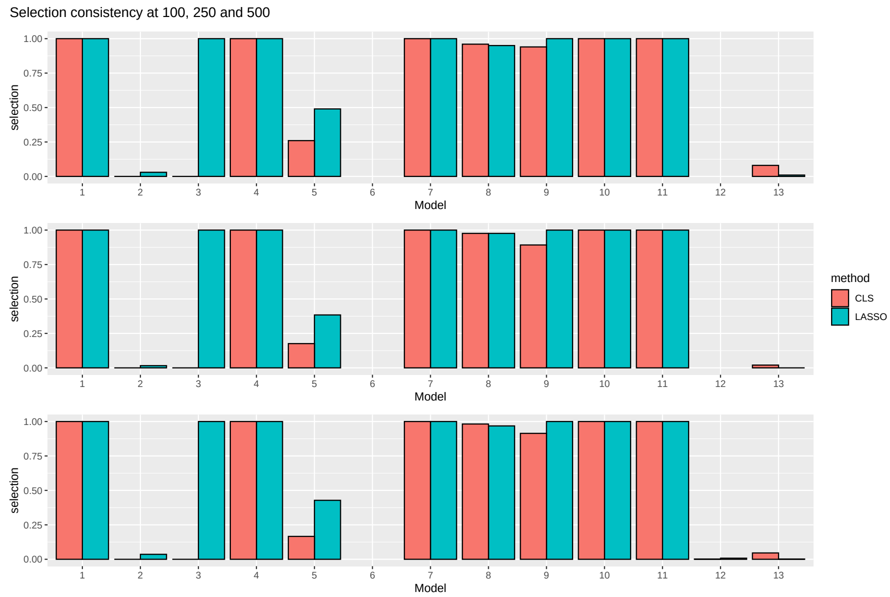

Figure 2. Model selection - large sample

4 CONCLUSIONS

This paper compared the accuracy of forecast combinations produced from selected individual forecasts versus forecast combinations where selection was not considered. A Monte Carlo experiment was used to assess the accuracy of five forecast combination methods across two different sized samples (small and large). Within each of these two samples, there were multiple replications (100, 250 and 500). The results indicate that selection improves the accuracy of forecast combinations based on SA and BG methods. However, the accuracy of forecasts based on regression based combinations (OLS and LAD) does not change significantly.

ACKNOWLEDGEMENT

The authors would like to thank the financial support received from the School of Accounting, Economics and Finance, Curtin University.

REFERENCES

Bates, J. M. & Granger, C. W. (1969), 'The combination of forecasts', *Operational Research Quarterly* **20**(4), 451–468.

Chan, F., Harris, M. N., Singh, R. B. & Yeo, W. B. E. (2022), *Nonlinear Econometric Models with Machine Learning*, Springer International Publishing, Cham, pp. 41–78.

Chan, F. & Mátyás, L. (2022), *Linear Econometric Models with Machine Learning*, Springer International Publishing, Cham, pp. 1–39.

Chan, F. & Pauwels, L. L. (2018), 'Some theoretical results on forecast combinations', *International Journal of Forecasting* **34**(1), 64–74.

Granger, C. W. & Ramanathan, R. (1984), 'Improved forecasting in time series using weighted combining', *Journal of Forecasting* **3**(2), 197–204.

Nowotarski, J., Raviv, E., Trück, S. & Weron, R. (2014), 'An empirical comparison of alternative schemes for combining electricity spot price forecasts', *Energy Economics* **46**, 395–412.

R Core Team (2021), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288.

Weiss, C. E., Roetzer, G. R. & Raviv, E. (2018), *ForecastComb: Forecast Combination Methods*. R package version 1.3.1.