Research Program on Forecasting

Can a subset of forecasters beat the simple average in the SPF?

Constantin Bürgi
The George Washington University
cburgi@gwu.edu

RPF Working Paper No. 2015-001
http://www.gwu.edu/~forcpgm/2015-001.pdf

March 27, 2015

RESEARCH PROGRAM ON FORECASTING
Center of Economic Research
Department of Economics
The George Washington University
Washington, DC 20052
http://www.gwu.edu/~forcpgm

Can a subset of forecasters beat the simple average in the SPF? *

Constantin Bürgi †

March 27, 2015

Abstract

The forecast combination literature has optimal combination methods, however, empirical studies have shown that the simple average is notoriously difficult to improve upon. This paper introduces a novel way to choose a subset of forecasters who might have specialized knowledge to improve upon the simple average over all forecasters in the SPF. In particular, taking the average of forecasters that recently beat the simple average more than the calibrated threshold of 52.5% of times can statistically significantly outperform the simple average for 10-year treasury bond yields, CPI inflation and unemployment at some horizons.

JEL: C22, C52, C53

Keywords: Forecast combination; Forecast evaluation; Multiple model comparisons; Real-time data; Survey of Professional Forecasters

---

*I would like to thank Neil Ericsson, Tara Sinclair, Herman Stekler and Benjamin Williams for their valuable comments and support.

† *cburgi@gwu.edu, The George Washington University, 2115 G Street, NW, # 374, Washington, DC 20052*

1 Introduction

Since Bates and Granger (1969), it has become well established that weighted combinations of forecasts perform better than individual forecasts. However, empirical studies summarized by Clemen (1989) or Timmermann (2006) showed several drawbacks from using optimal weights. In particular, it appears to be quite difficult to improve upon the simple average of individual forecasts, which is often explained by the estimation of optimal weights. In addition, other combination methods like past performance do not perform very well or select the best forecasters who might have specialized knowledge. A more recent study of the variables in the ECB Survey of Professional Forecasters by Genre et al. (2013) found similar results using a wide array of combination methods.

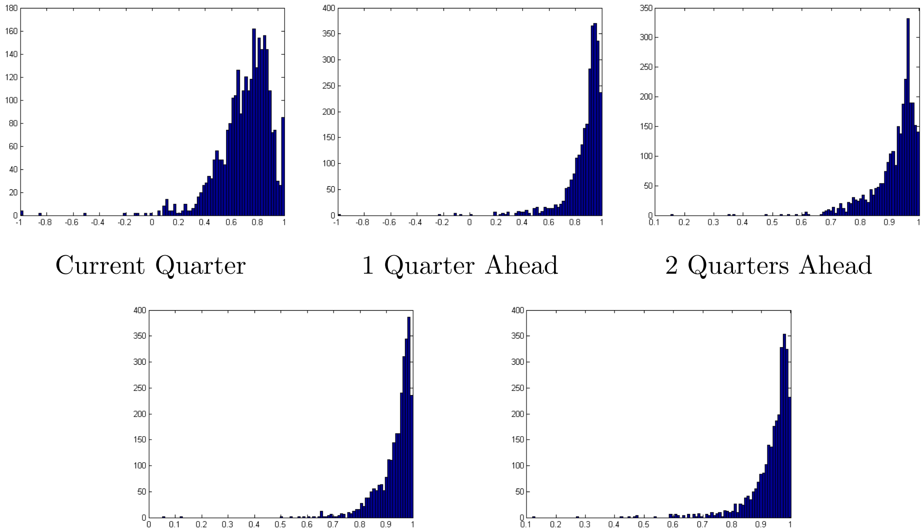
Due to these findings, it is not surprising that many surveys collecting forecasts report the simple average of forecasts as the benchmark. However, as Blix et al. (2001) showed for Consensus Economics forecasts, there are forecasters that beat this simple average even over extensive periods based on mean squared errors (MSE).

The existence of individual forecasters beating the simple average based on MSE immediately leads to the question, why it is this difficult to find the best forecasters and to improve upon the simple average. This paper will show that this issue arises as well in the Survey of Professional Forecasters (SPF) conducted by the Philadelphia Fed and links it to the correlation among forecasters. High correlation among forecast errors can lead to many individual forecasters outperforming the simple average by chance. A new method to obtain better forecasters with potential specialized knowledge will be introduced. This new method is subsequently applied to CPI, unemployment and bond yield forecasts in the SPF. It will be shown that using this method statistically significant improvements upon the simple average can be obtained for several forecasts.

2 Model

In the SPF, a similar pattern to the one found by Blix et al. (2001) emerges across variables. For example, of the CPI inflation forecasters with at least 20 forecasts in the period 1992:Q1-2013:Q2, around 20% have lower MSEs than the simple average at every horizon. At the same time, the histograms of the pairwise correlation matrix of SPF CPI forecasters in the period 1992:Q1-2013:Q2 shown in Figure 1 show the distribution is very skewed towards high correlations for all five horizons. The average pairwise correlation for current quarter forecasters is 0.65 and around 0.8-0.85 for the other horizons.

Figure 1: Distribution of correlations among CPI forecasters



Current Quarter | 1 Quarter Ahead | 2 Quarters Ahead

3 Quarters Ahead | 4 Quarters Ahead

To show that the out performance of individual forecasters is linked to the high correlation, assume that forecast errors have the form

$$\nu_{it} = \delta\gamma_t + (1 - \delta)\varepsilon_{it}, \tag{1}$$

where the individual forecast errors are the weighted sum of a common forecast error γ_t and an idiosyncratic error ε_{it}, which both are iid $(0, \sigma^2 < \infty)$. By construction, variance of ν_{it} for $t \to \infty$ is $\delta^2\sigma^2 + (1-\delta)^2\sigma^2$, while the asymptotic variance of the simple average is $\delta^2\sigma^2 < \delta^2\sigma^2 + (1 - \delta)^2\sigma^2$.

While there are never individuals that beat the simple average for any correlation $\rho < 1$ in infinite samples, this does not hold in finite samples due to the Law of Large Numbers. As the gains from averaging are smaller at high correlations (high δ), the percentage of forecasters beating the simple average become larger.

To identify the forecasters with private knowledge that are indeed better than other forecasters, it is necessary to find a selection criterion that yields less false positives than measures based on MSE or mean absolute error (MAE) at high correlations. One such alternative real-time method to obtain the best forecasters is based on the past performance rank, similar to Stekler (1987) and Batchelor (1990). In particular, a subset of forecasters is created by choosing forecasters that have already made at least n forecasts and have beaten the simple average more often than a certain percentage threshold p. As an alternative to this rule based threshold approach, one could also use an estimated approach like impulse indicator saturation as described in Ericsson and Reisman (2012).

While it might be desirable to increase the threshold until the percentage of individual forecasters beating it by chance reaches a very small number, this would also decrease the probability of detecting better forecasters.

3 Empirical application

This alternative method of picking the best forecasters will be tested on three variables CPI, Unemployment rate and 10-year treasury bond yield from the SPF as they are not subject to large revisions, unlike GDP. Also, D'Agostino et al. (2012) showed using a similar method that there is little evidence of better forecasters in GDP. Bond yield forecasts only start in 1992, which is the date chosen for all three series. The sample ends in Q1 2013 and includes the 2008 crisis. The subset is chosen based on the method described above, with $n = 10$. To ensure that the subset includes a hand full of forecasters every period, the threshold is set at $p = 52.5\%$.

The overall root MSE (RMSE) of the average of the subset of forecasters is compared to the RMSE of the simple average. Table 1 shows the percentage improvement of the subset relative to the overall average from the current quarter (cq) forecast [1] to the four quarter ahead (4q) forecast. Negative signs imply an improvement and the Diebold and Mariano (1995) test with quadratic loss function and h=1-5 and the adjustment by Harvey et al. (1997) (DM-test) is used to determine the significance.

The single largest gain is found in current quarter CPI forecasts where the RMSE of the subset with specialized knowledge improves by 26.74% over the SPF average. For the CPI, there is no gain relative to the overall SPF average for other horizons.

For unemployment, there is a gain at most horizons, however only 2 quarters ahead forecast are significant at the 5% level and current quarter forecast at the 10% level.

10-year government bond yields show most significant gains of all three variables, which could stem from specialized knowledge of some forecaster. In particular at the very short horizon and the longer term, the subset outperforms the overall average, while the gains are less significant for medium horizons.

In addition to comparing MSE of the two methods, it is important to check, if other measures lead to similar results. Table 2 show the relative gains of

[1] Forecasts for the SPF are collected in the middle month of a quarter

Table 1: Percentage RMSE improvement relative to simple average

	CPI	Unemployment	10-yr Treasury
cq	-26.74*	-5.29^{+}	-13.39**
1q	0.73	-4.05	-3.61^{+}
2q	3.01	-6.71*	-2.68
3q	-0.89	-4.80	-7.71*
4q	1.11	1.65	-4.12*

$^{+}$ significant improvement at 10% level, * at 5% level and ** at 1% level based on one sided DM-test.

Table 2: Percentage MAE improvement relative to simple average

	CPI	Unemployment	10-yr Treasury
cq	-22.04**	-4.36	-13.67**
1q	1.18	-3.08	-6.04**
2q	4.78	-13.50*	-2.04
3q	-0.76	-4.50	-11.09**
4q	-0.58	6.19	-4.92*

$^{+}$ significant improvement at 10% level, * at 5% level and ** at 1% level based on one sided DM-test.

the subset of forecasts based on mean absolute error (MAE) and Table 3 what percentage of periods the subset of best forecasters beats the simple average. The results are quite similar to MSEs, as current quarter inflation and bond yield forecasts are highly significant as well as the longer term forecast for bond yields. Most other forecasts remain insignificantly different from the simple average.

Other important robustness checks include the sensitivity towards different values of the percentage threshold p, the minimum number of forecasts required n and the performance over different time periods. In this section, the sensitivity analysis for bond yields is show and the sensitivity analysis for the other variables can be found in the appendix. As stricter thresholds might lead to periods without any forecaster in the subset, the simple average replaces the subset for those periods.

Figure 3 shows a clear percentage improvement of the subset relative to the simple average based on RMSE over different threshold values p. Figure 3 shows

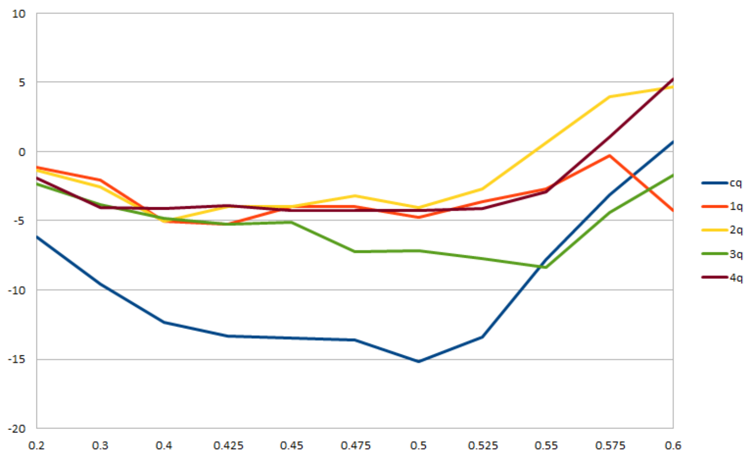
Table 3: Share of forecasts that beat the simple average

	CPI	Unemployment	10-yr Treasury
cq	0.68**	0.53	0.64**
1q	0.47	0.58+	0.63*
2q	0.37	0.64**	0.78**
3q	0.54	0.53	0.74**
4q	0.57	0.41	0.66**

+ significant at 10% level, * at 5% level and ** at 1% level based on a one sided test.

that the significance of these gains is broadly stable across values of p, with the 10%, 5% and 1% threshold highlighted.

Figure 2: % RMSE improvement for different values of p



For different values of n, a similar pattern emerges, as Figures 3 and 3 show.

The gains for bond yields are persistent over time as figure 3 shows. In this graph, the subset tends to be closer to the actual number than the simple average (positive value). While there are some periods, where the simple average tends to perform better, there is no clear cyclical pattern.

As another robustness check, do forecasters with private knowledge that predict well at one horizon, also perform well for other horizons? In particular, if one forecaster has private information that renders his 1 quarter ahead forecasts among the best, that information is likely to be very valuable for 2q and 3q ahead forecasts as well he should perform very well at other horizons as well.

Figure 3: DM-stat for different values of p

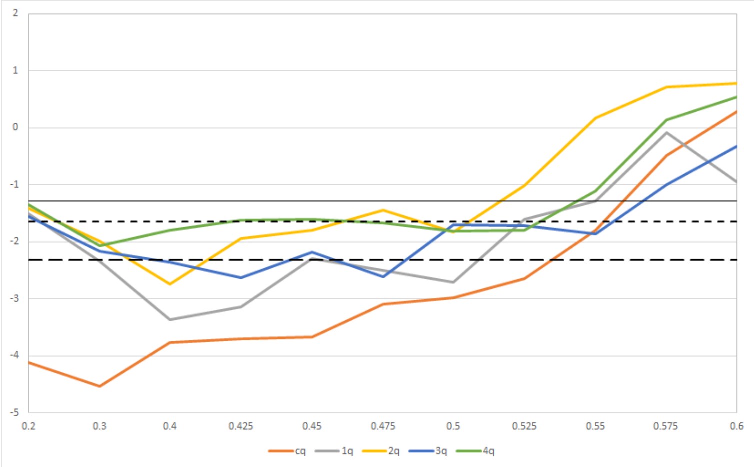


Figure 4: % RMSE improvement for different values of n

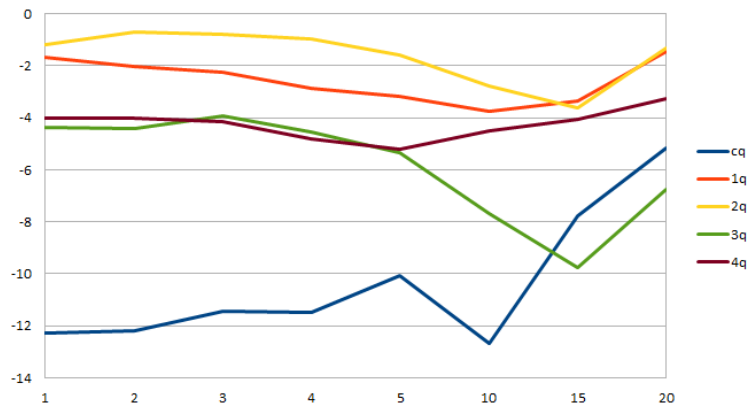


Figure 5: DM-stat for different values of n

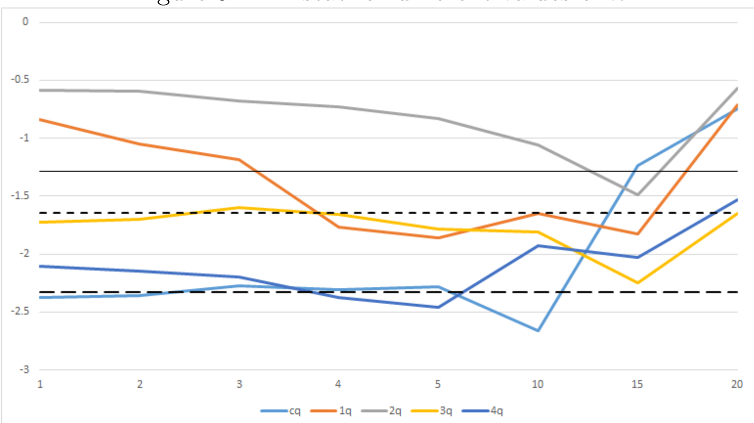
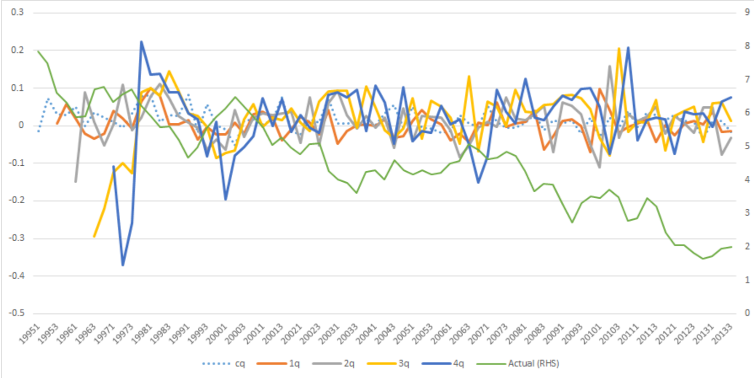


Figure 6: Difference in forecast errors between simple average and subset



As half of the quarter has already passed, forecasting the current quarter might be different from forecasting other quarters.

Table 4: Percentage RMSE improvement relative to simple average, using the best 1 quarter ahead forecasters

	CPI	Unemployment	10-yr Treasury
cq	17.11	-7.54	-1.90^{+}
2q	-0.07	-9.27^{+}	-4.76**
3q	1.90	-9.51	-5.06*
4q	0.55	-8.89	-5.94*

$^{+}$ significant improvement at 10% level, * at 5% level and ** at 1% level based on one sided DM-test.

As Table 4 shows, forecasters that perform well 1 quarter ahead tend to perform better at other horizons as well for bond yields and based on MSE. This could hint at specialized knowledge for bond forecasting, but much less so for unemployment and CPI.

4 Conclusion

It was shown that selecting the best forecasters based on MSE or MAE might not always lead to the best forecasters, if the correlation among their forecast errors is high.

Subsequently a new method to obtain a subset of best forecasters was introduced and it was shown that this method is able to improve statistically significantly over the overall SPF average for 10-year government bond yields, CPI and unemployment for some horizons. This results holds across several forecast evaluation methods. However, it does not show significant improvement across all horizons. In addition, there is some evidence that forecasters that are in the subset for one horizon tend to do better on other horizons as well.

Further research might be able to determine, if it is indeed specialized knowledge why there appear to be gains across horizons in bond yields, but much more limited gains for CPI and unemployment.

References

Batchelor, R. (1990). All forecasters are equal. *Journal of Business & Economic Statistics*, 8(1):143–44.

Bates, J. and Granger, C. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468.

Blix, M., Wadefjord, J., Wienecke, U., and Adahl, M. (2001). How good is the forecasting performance of major institutions? *Sveriges riksbank economic review*, pages 38–68.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, 5(4):559–583.

D'Agostino, A., Mcquinn, K., and Whelan, K. (2012). Are some forecasters really better than others? *Journal of Money, Credit and Banking*, 44(4):715–732.

Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.

Ericsson, N. and Reisman, E. (2012). Evaluating a Global Vector Autoregression for Forecasting. *International Advances in Economic Research*, 18(3):247–258.

Genre, V., Kenny, G., Meyler, A., and Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting*, 29(1):108–121.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2):281–291.

Stekler, H. O. (1987). Who forecasts better? *Journal of Business & Economic Statistics*, 5(1):pp. 155–158.

Timmermann, A. (2006). Chapter 4 forecast combinations. In G. Elliott, C. G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1 of *Handbook of Economic Forecasting*, pages 135 – 196. Elsevier.

A Graphs CPI

Figure 7: % RMSE improvement for different values of p

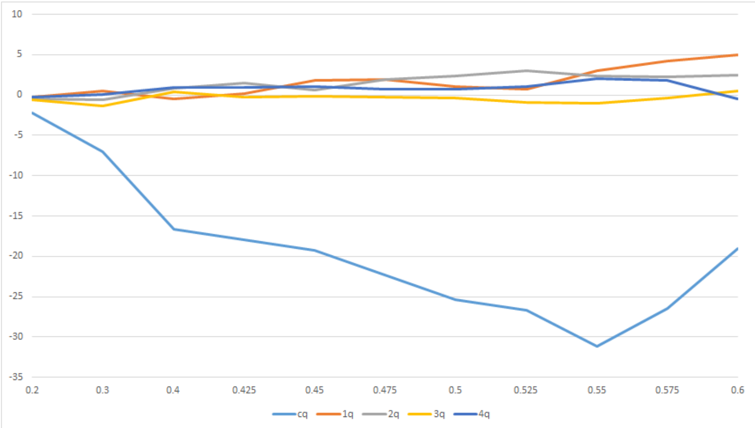


Figure 8: DM-stat for different values of p

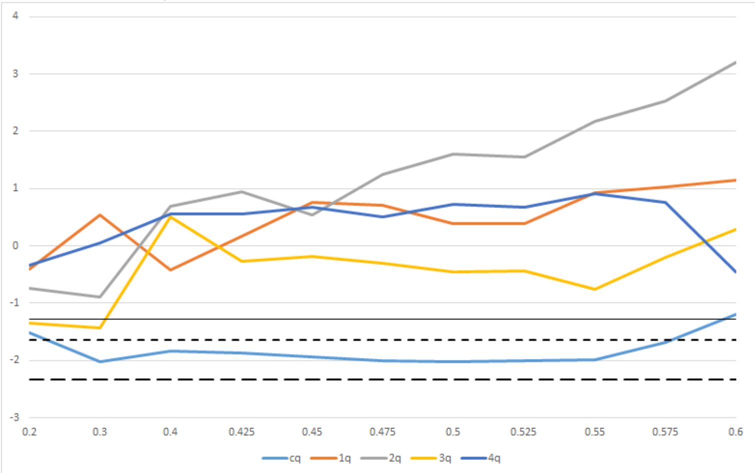


Figure 9: % RMSE improvement for different values of n

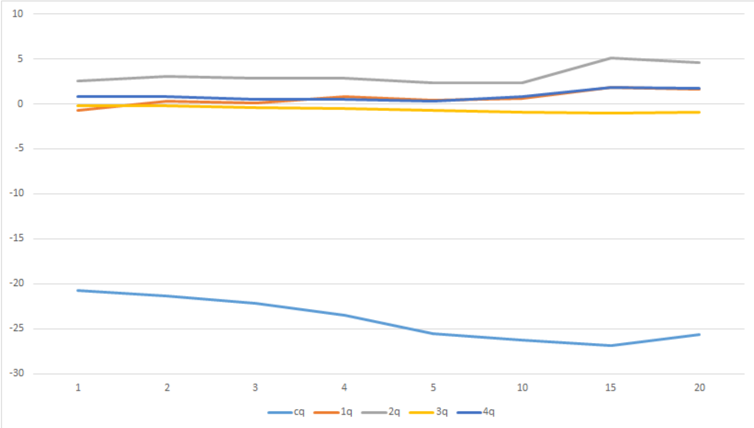


Figure 10: DM-stat for different values of n

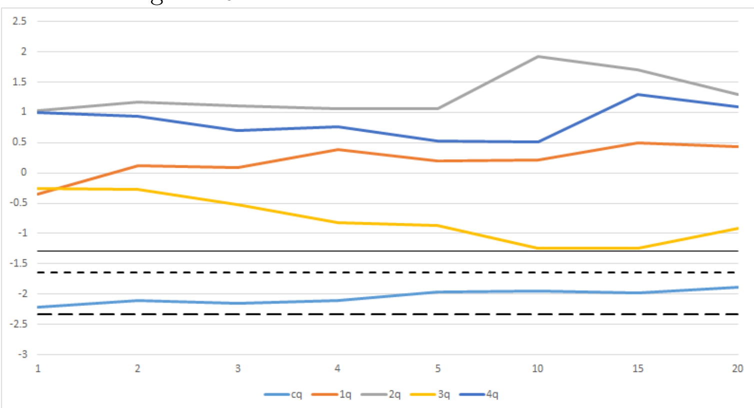
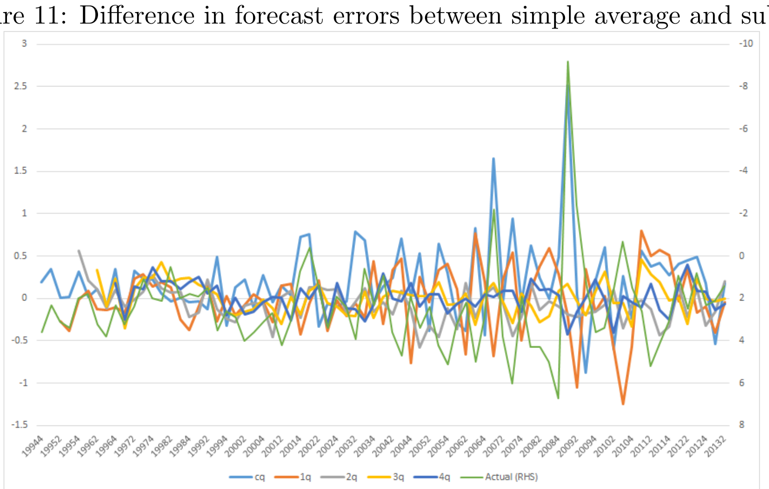


Figure 11: Difference in forecast errors between simple average and subset



B Graphs Unemployment

Figure 12: % RMSE improvement for different values of p

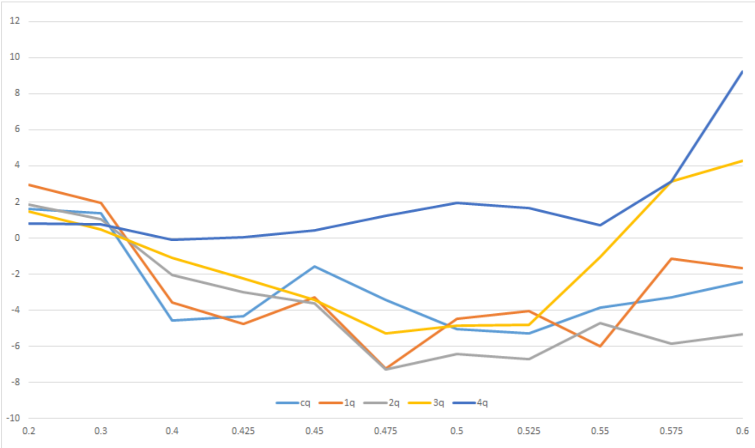


Figure 13: DM-stat for different values of p

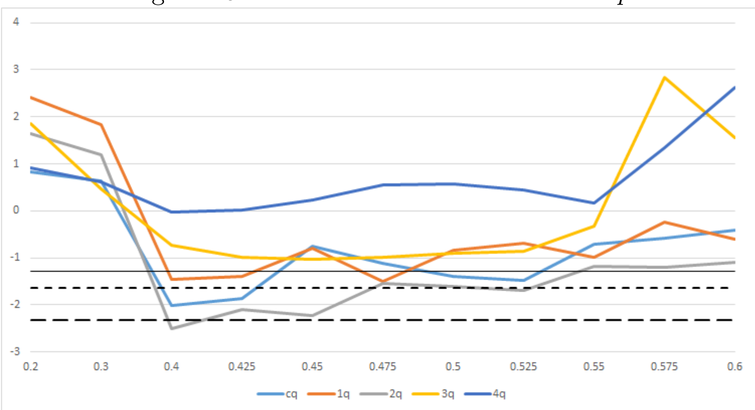


Figure 14: % RMSE improvement for different values of n

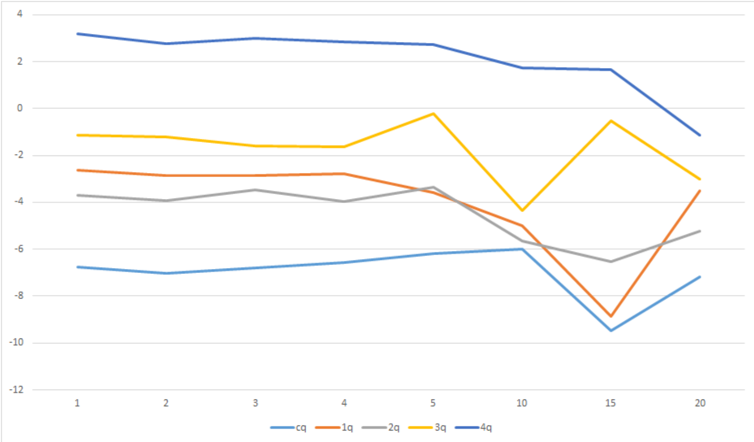


Figure 15: DM-stat for different values of n

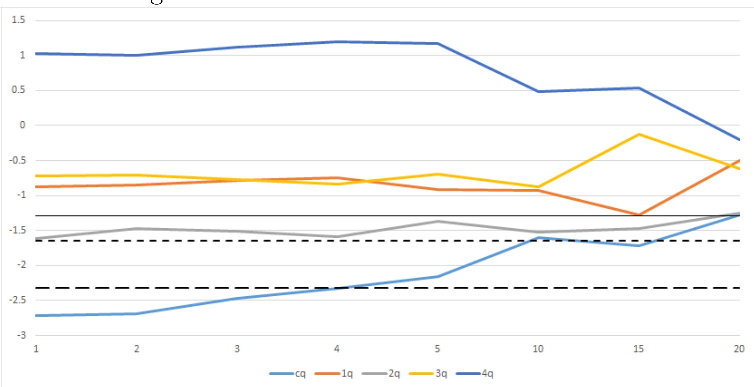


Figure 16: Difference in forecast errors between simple average and subset