Multivariate Rank-Based Forecast Combining Techniques

Matthias Klapper

Department of Statistics, University of Dortmund,

44221 Dortmund, Germany

klapper@amadeus.statistik.uni-dortmund.de

Abstract: We analyze macroeconomic data using univariate and multivariate forecast combining techniques. We simulate forecast errors with different variance-covariance structures. The simulations are used to compare the performance of univariate and multivariate combining techniques.

Keywords: Variance-covariance structure, simulation, combination of forecasts, multivariate combination of forecasts.

Acknowledgement: Supported by Deutsche Forschungsgemeinschaft, Graduiertenkolleg „Angewandte Statistik“.

AMS 1991 Subject Classification: 62G30

1. Introduction

Previous research showed that if we combine several forecasts of the same event we get a combined forecast that is better than the individual forecasts or a simple average, e.g. Newbold and Granger (1974), Russel and Adam (1987), Klapper (1998a), and Klapper (1998b) to name a few. Usually several variables are forecasted at the same time e.g. economic indicators, indices of stock markets or election results of several parties. In the past, each of these variables has been combined separately and only Wenzel (1998) has done some steps towards a multivariate approach. We will extend the rank-based forecast combining techniques into a

multivariate framework by using information of past performance of several variables for determining the weights of one variable.
Chapter 2 will present the multivariate combining techniques and apply them to the German macro-economic forecast data. We will then use simulations in Chapter 3 to look at the combination of variables with different variance-covariance structures. Chapter 4 concludes with a summary of the results.

2. Multivariate Combination Techniques

The idea of multivariate forecast combinations is that there is information about the quality of future forecasts hidden in the past performance of forecasts for other variables. An explanation of the incremental new information from a second variable could be that other researchers with more experience could be involved with the other forecast. The multivariate techniques considered are based on this assumption. Some methods are using the correlations between variables to determine the weights and are therefore considering the covariance-structure of past peformances. Past forecast errors from variables that are more correlated with variable A are getting more weight towards the determination of the weight coefficients than variables that are less correlated with variable A. Other methods do not include the correlation aspect.
We will use the following multivariate rank-based combining techniques and compare their performance to univariate combining techniques.

<u>Multivariate combining methods</u>:

- To determine the weights for the different forecaster for one variable, the method All1/k takes the inverse sum of ranks of all variables for the last 10 time points for each forecaster divided by the sum of the inverses of all forecasters. This yields coefficients for each forecaster that are greater than zero and add up to one, which is the same for all other rank based methods. This method is giving all variables equal weights.
- Allo1/2k works like All1/k but weights the ranks of the other variables only with 1/2. This creates different weights for the variables.
- Corr weights the ranks for each variable with the correlation of that variable with the variable whose weights are calculated.
- Corr2 works like Corr but uses squared correlations. Other variables get lower weights than with Corr.
- Corr0.5 is giving variables a weight of 1 if the correlation is greater or equal 0.5. This favors higher and excludes smaller correlations.

<u>Univariate combining methods</u>:

- The method called Rank takes the inverse of the sum of the ranks of the last 10 time points for each forecaster divided by the sum of the inverses of all forecasters.
- The QRank technique does the same as Rank but takes the quadrupled ranks instead of the simple ranks.
- The cmse technique as explained in Russell and Adam (1987) takes the inverse of the MSE of the past 10 performance periods for each forecaster and puts it into relation to the sum of the inverse MSEs of all forecasters.
- The New/Gr method explained in Newbold and Granger (1974) takes the row sums of the inverse covariance matrix of the forecast errors and divides it by the sum of all elements of the inverse covariance matrix. This method is equivalent to an OLS regression approach.

Applying these methods to the macro-economic data presented in Klapper (1998a) we get the RMSEs displayed in Table 1. All techniques yield a lower RMSE than the simple average except the New/Gr method and QRank for Private Consumption and All1/k for Consumer prices. QRank has the lowest RMSE for GDP and Public Consumption and cmse performs best for Export, Import, and Consumer Prices. For these five variables the univariate combination methods score better than the multivariate techniques. For Private Consumption, the multivariate outperform the univariate techniques with Corr0.5 having the best result. This method also performs best among the multivariate techniques for Public Consumption. Other multivariate techniques like Allo1/2k for GDP and Export, Corr for Import, and Corr2 for Consumer Prices perform well among the multivariate techniques and are able to beat the simple average. Therefore, we will use a simulation study to examine the impact of the variance-covariance structure on the performance of these methods. For an analysis of the multivariate variance-covariance structure see Klapper (1998b).

Table 1 - Univariate vs. Multivariate Techniques

	Technique	GDP	Priv Cons	Publ Cons	Export	Import	Cons Pr
Uni-	Rank	0,994	0,996	0,975	0,996	0,998	0,995
variate:	QRank	0,986	1,000	0,965	0,999	0,994	0,998
	cmse	0,990	0,996	0,969	0,993	0,991	0,991
	New/Gr	2,623	1,736	2,835	2,153	1,305	1,874
Multi-	All 1/6	0,996	0,995	0,989	0,998	0,998	1,002
variate:	all other 1/10	0,995	0,995	0,983	0,997	0,998	0,999
	corr	0,997	0,994	0,984	0,998	0,998	0,997
	corr^2	0,997	0,995	0,981	0,998	0,999	0,996
	corr>0.5	0,998	0,992	0,975	0,998	0,999	0,997

Rounded RMSEs in relation to the simple average's RMSE.

3. Simulation

To keep things simple we take only two variables with three forecasters and assume the errors to be normally distributed with no autocorrelation present. We will vary the correlations within the variables between the forecasters (corin) and the correlations between the variables within the forecasters (corex).

The first three columns of the covariance matrix represent the first variable, one for each of the three forecaster, the other three for the second variable. With i for within-variable correlation (corin) and e for between-variable correlation (corex) the correlation matrix looks like this:

$$\text{Cor} = \left(\begin{array}{ccc|ccc} 1 & i & i & e & e & e \\ i & 1 & i & e & e & e \\ i & i & 1 & e & e & e \\ \hline e & e & e & 1 & i & i \\ e & e & e & i & 1 & i \\ e & e & e & i & i & 1 \end{array} \right)$$

We will use the levels of 0.1, 0.3, 0.5, and 0.85 for corin and 0.1, 0.3, 0.5, 0.8, 0.9, and 0.95 for corex. The variances of the three forecaster's forecast errors are assumed to be constant over time at 2, 1.5, and 1 for both variables. We will use the first variable with the univariate techniques and as the main variable for the multivariate methods. The influence of the second variable will be used for the multivariate aspect. Each time series has 21 time points to be close to the real data. We will use 1,000 simulation runs for each of the 24 scenarios.
Both, Table 2 and Figure 1 show the percentage of times the RMSE is lower than the simple average's RMSE. The graphs on the left hand side show all combining techniques tested, the graphs on the right only the multivariate techniques. This was done to increase the scale for better visibility of differences. All multivariate combining techniques have at least 50% of the time a lower RMSE than the simple average with some methods getting up to more than 90% for high corin. The multivariate techniques beat the simple average usually more often than the univariate methods. For lower corex of 0.1 and 0.3, All1/k performs best, beating the second-best method by up to 3 percentage points, e.g. Allo1/2k for 0.1 corin. For higher corex the correlation based methods get better but do not significantly outperform All1/k. The highest margin is between All1/k and corr for 0.3 corin and 0.5 corex. This indicates that using all variable's ranks of forecast errors improves the estimation of different variances especially for these short time series of only 21 observations. The univariate techniques that are supposed to be invariant towards change of corex by constant corin are showing partially big differences of e.g. 8.3 percentage points for QRank at 0.1 corin between 0.5 and 0.95 corex., where all univariate methods show big differences. This could be due to chance associated with the different simulation runs.
From Table 3 and Figure 2 we can see the average RMSE as percentage of the simple average's RMSE. For better visibility again the graphs on the left hand side show all combining techniques tested, the graphs on the right only the multivariate techniques. For all scenarios the multivariate combining techniques yield between 0.6% and 1.4% reduction in variability. In general it is valid, that the higher the within-variable correlation, the better the performance since predictability increases. If we increase the between-variable correlation, all methods perform worse. The graphs on the right show that the multivariate methods are very close to each other regardless of change of the within-variable and between-variable correlations.
For a high within-variable correlation of 0.85 the New/Gr method has the lowest average RMSE except for 0.5 corex. At a corin of 0.3 the univariate rank-based methods and cmse perform best, and if corin is 0.3 or lower the multivariate techniqes lead to the best results. All1/k and Allo1/2k are better for corex of 0.1-0.5, the correlation-based methods corr, corr2, and corr0.5 score better for higher corex of 0.8-0.95.

4. Summary

We were able to develop multivariate forecast combining techniques that improve over univariate techniques for certain variance-covariance structures. For the German macro-economic data, improvements of the RMSE over the simple average range beween 0.1 and 2.5 percentage points. For Private Consumption, multivariate outperform univariate techniques decreasing the RMSE by 0.4%.
The simulation study showed that multivariate methods perform best if there is low correlation between the forecasters. In this case, they yield a 1-3% improvement in mean RMSE over the univariate methods and beat the simple average 2-10 percentage points more often. Multivariate methods beat the simple average for low between-variable correlation by 2.1-10.6

percentage points and are more robust in their performance towards changes in the variance-covariance structure.

Table 2 - % of Times Simple Average is Beaten

corin	Corin	Rank	QRank	cmse	New/Gr	All1/k	Allo1/2k	corr	corr2	corr>0.5
0.85	0.10	84	89	90	58	92	90	88	86	86
0.85	0.30	81	84	86	60	89	88	88	85	84
0.85	0.50	82	86	87	56	90	89	90	89	89
0.85	0.80	84	88	90	58	91	89	91	91	91
0.85	0.90	82	86	89	58	89	87	89	89	89
0.85	0.95	82	85	86	57	89	88	88	88	89
0.50	0.10	63	58	61	34	71	69	68	67	67
0.50	0.30	62	61	64	34	72	70	71	70	69
0.50	0.50	64	60	63	34	71	69	69	67	67
0.50	0.80	62	60	60	30	69	67	68	68	69
0.50	0.90	59	57	58	30	65	64	66	66	65
0.50	0.95	61	59	59	34	69	66	69	69	69
0.30	0.10	58	51	52	30	64	64	62	61	61
0.30	0.30	57	51	52	30	68	64	67	64	62
0.30	0.50	56	50	50	30	64	61	65	64	64
0.30	0.80	52	50	52	27	60	60	61	62	60
0.30	0.90	55	49	49	29	62	59	62	62	62
0.30	0.95	55	50	50	29	64	62	64	64	64
0.10	0.10	54	45	45	27	62	59	58	57	57
0.10	0.30	56	43	45	29	66	63	66	64	62
0.10	0.50	50	40	41	25	56	55	57	56	56
0.10	0.80	51	43	43	23	56	54	57	58	56
0.10	0.90	55	47	48	25	59	59	59	59	59
0.10	0.95	53	48	49	26	59	58	59	59	59

Numbers in rounded percentages.
corin = within-variables correlation, corex = between-variables correlation.

Table 3 - % less Mean RMSE than Simple Average

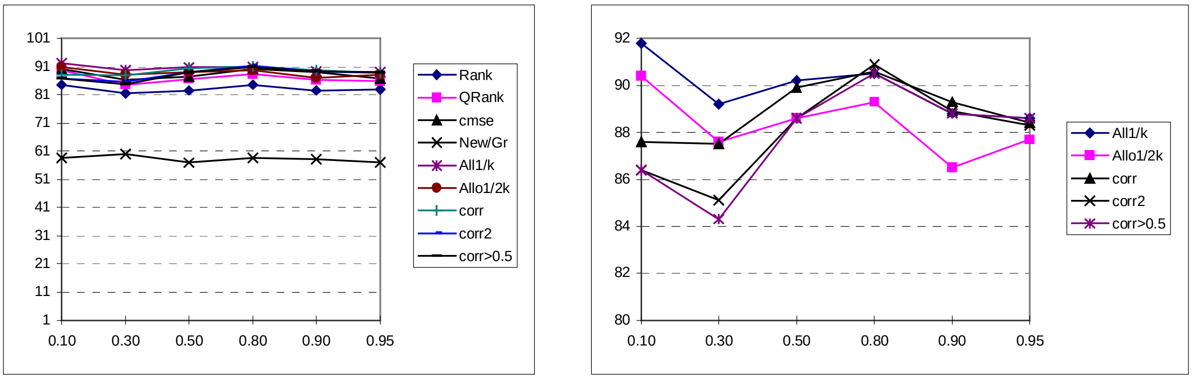
corin	Corin	Rank	QRank	cmse	New/Gr	All1/k	Allo1/2k	corr	corr2	corr>0.5
0.85	0.10	1,4	4,0	3,9	4,2	1,4	1,4	1,4	1,4	1,4
0.85	0.30	1,4	3,8	3,8	5,1	1,4	1,4	1,4	1,4	1,3
0.85	0.50	1,3	3,6	3,6	3,0	1,3	1,3	1,3	1,3	1,3
0.85	0.80	1,4	3,8	3,8	4,9	1,4	1,4	1,4	1,4	1,4
0.85	0.90	1,3	3,8	3,8	3,8	1,3	1,3	1,3	1,3	1,3
0.85	0.95	1,3	3,5	3,5	3,7	1,3	1,3	1,3	1,3	1,3
0.50	0.10	1,1	2,1	2,4	-7,7	1,3	1,3	1,2	1,2	1,2
0.50	0.30	1,2	2,4	2,7	-7,3	1,4	1,4	1,4	1,3	1,3
0.50	0.50	1,3	2,3	2,5	-7,5	1,3	1,3	1,3	1,3	1,3
0.50	0.80	0,8	1,6	1,6	-9,3	1,0	1,0	1,0	1,0	1,0
0.50	0.90	0,8	1,5	1,6	-9,6	0,9	0,8	0,9	0,9	0,9
0.50	0.95	0,9	2,0	2,0	-8,0	1,0	1,0	1,0	1,0	1,0
0.30	0.10	1,0	0,8	1,3	-9,1	1,3	1,2	1,2	1,2	1,2
0.30	0.30	1,0	0,9	1,4	-8,7	1,3	1,3	1,3	1,2	1,2
0.30	0.50	0,9	0,5	0,8	-10,6	1,2	1,2	1,2	1,1	1,1
0.30	0.80	0,6	0,5	0,8	-10,9	0,8	0,8	0,8	0,8	0,8
0.30	0.90	0,6	0,5	0,7	-10,9	0,7	0,7	0,7	0,7	0,7
0.30	0.95	0,6	0,7	0,8	-10,8	0,7	0,7	0,7	0,7	0,7
0.10	0.10	0,9	-1,3	-0,8	-10,2	1,4	1,3	1,1	1,0	1,0
0.10	0.30	1,2	-1,1	-0,8	-9,5	1,7	1,7	1,7	1,5	1,4
0.10	0.50	0,4	-1,8	-1,5	-11,9	0,8	0,7	0,8	0,7	0,7
0.10	0.80	0,4	-0,5	-0,4	-12,3	0,7	0,6	0,7	0,7	0,7
0.10	0.90	0,6	0,0	0,2	-11,6	0,7	0,7	0,7	0,7	0,7
0.10	0.95	0,4	-0,1	0,1	-11,8	0,6	0,6	0,6	0,6	0,6

Numbers in percentages that a technique's RMSE is lower than the simple average's RMSE.

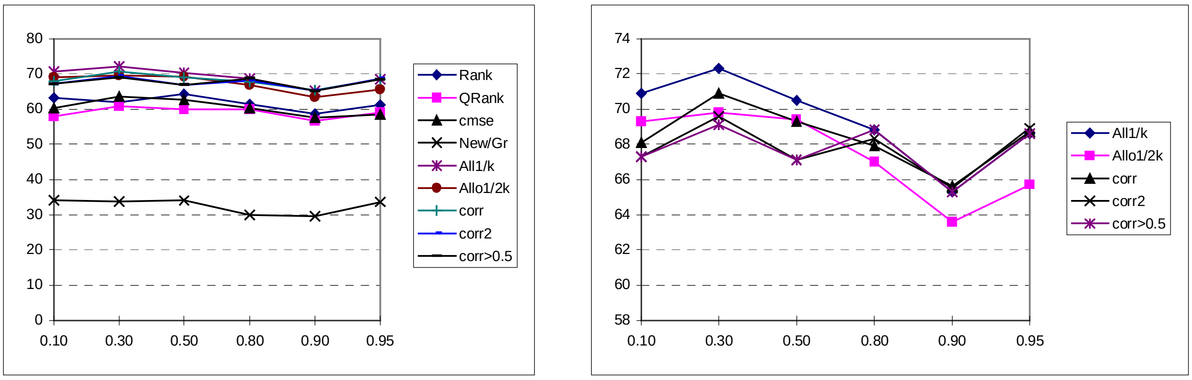
corin = within-variables correlation, corex = between-variables correlation.

Figure 1 - Multivariate Simulation: % of time SA is beaten

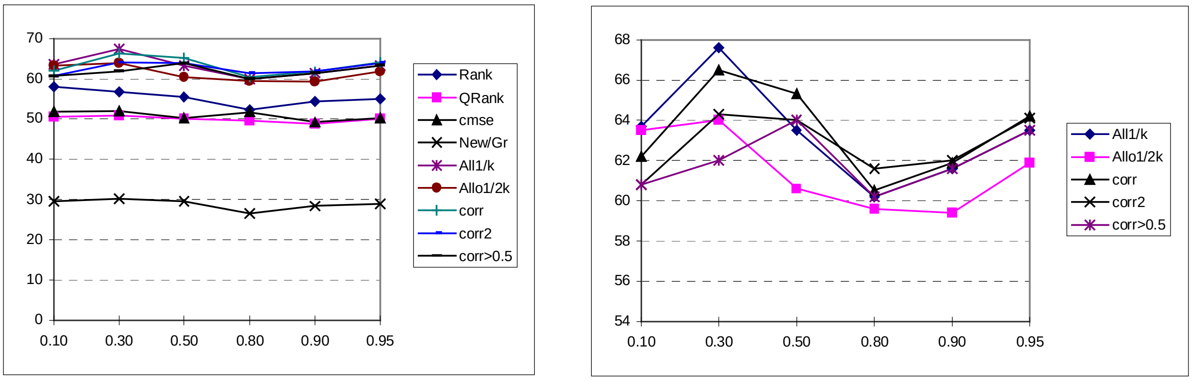
Corin = 0.85



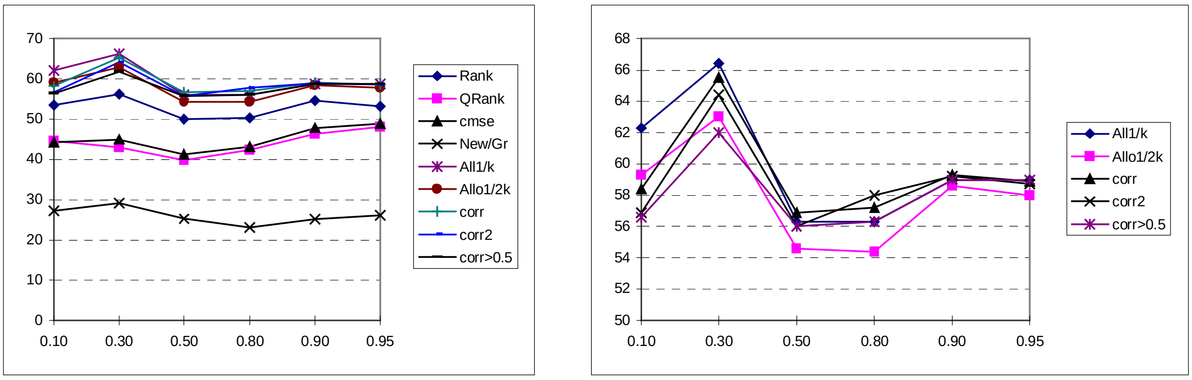
Corin = 0.50



Corin = 0.30



Corin = 0.10

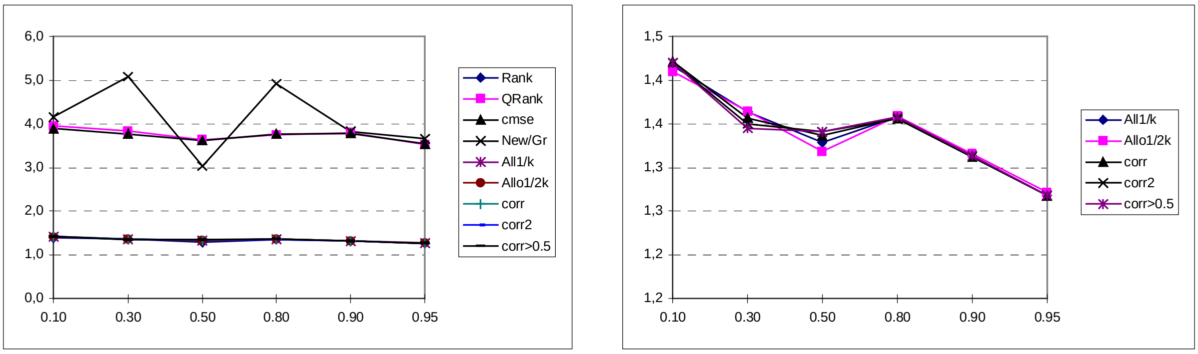


Corin = Correlation within variables (between forecasters).
x-axes: correlation between variables (corex), y-axes: %, not to scale.

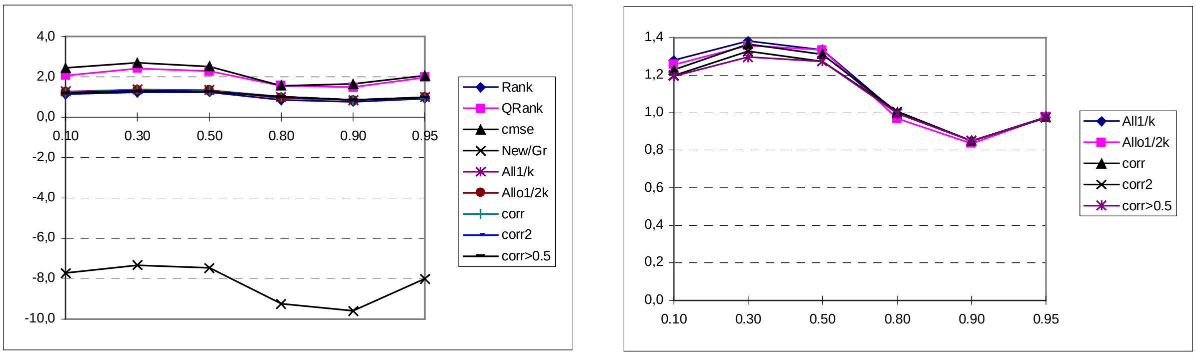
Left: All combining methods, right: only multivariate techniques.

Figure 2 - Multivariate Simulation: % less RMSE than SA

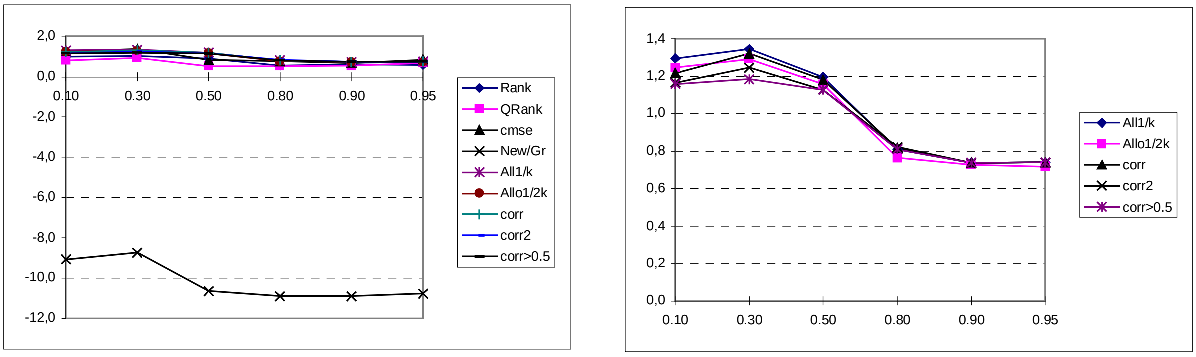
Corin = 0.85



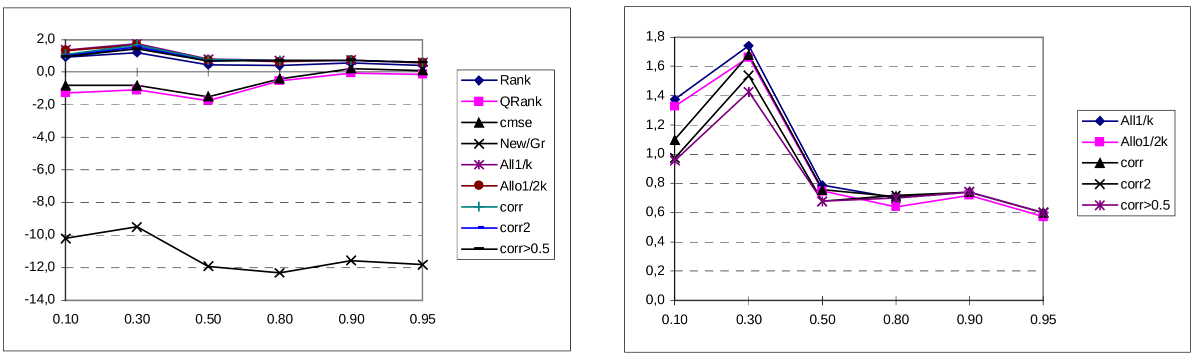
Corin = 0.50



Corin = 0.30



Corin = 0.10



Corin = Correlation within variables (between forecasters).
x-axes: correlation between variables, y-axes: %, not to scale.
Left: All combining methods, right: only multivariate techniques.

Bibliography

- Klapper, Matthias (1998a): Combining German Macro Economic Forecasts using Rank-Based Techniques, Technical Report 19/1998, SFB 475, University of Dortmund
- Klapper, Matthias (1998b): The Influence of the Variance-Covariance Structure on the Performance of Forecast Combining Techniques, Technical Report 39/1998, SFB 475, University of Dortmund
- Newbold, P. and C.W.J. Granger (1974): Experience with Forecasting Univariate Time Series and the Combination of Forecasts, J.R. Statist. Soc. A Vol. 137, 131-165
- Russell, Thomas D. and Everett E. Adam, Jr. (1987): An Empirical Evaluation of Alternative Forecast Combinations, Management Science Vol. 33, 1267-1276
- Wenzel, Thomas (1998): Pitman-closeness and the linear combination of multivariate forecasts, Technical Report 34/1998, SFB 475, University of Dortmund