# The Influence of the Variance-Covariance Structure on the Performance of Forecast Combining Techniques

Matthias Klapper
*Department of Statistics, University of Dortmund,*
*44221 Dortmund, Germany*
*klapper@amadeus.statistik.uni-dortmund.de*

**Abstract:** We simulate forecast errors with different variance-covariance structures based on macroeconomic data. The simulations are used to compare the performance of different forecast combining techniques.

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

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

**AMS 1991 Subject Classification: 62G30**

## 1. Introduction

This paper will be an extension of the analysis of the German macro economic data presented in Klapper (1998) that should be consulted for a detailed description of the data. While the primal focus of Technical Report 19/1998 was more on presenting and describing the data and proposing rank based methods for forecast combination, this paper will be more concerned with a thorough data analysis. The goal will be to give an answer to the question when and why a combining technique is superior to another.
We will first analyze the variance-covariance structure of the given macro economic data set in Chapter 2. This outcome will be used in a simulation study with the results presented in

Chapter 3. The concluding Chapter 4 contains a summary and an outlook for future research. Appendix A lists detailed formulas.

# 2. Analysis of the Variance-Covariance Structure

Our data set contains 6 variables having very similar variance-covariance structures. As displayed in Table 1 the relationship of highest to lowest variances of the 7 forecasters range from almost 3:1 for public consumption to 1.5:1 for exports and consumer prices. In all cases there are several forecasters with very similar variances giving variance based methods a hard time to determine weights, e.g. for the exports that have 4 forecasters with very similar variances at a maximum difference of 8%. These relationships lead standard tests for equality of variances like Bartlett’s test come to the conclusion that there are no differences. Table 1 shows the p-values for rejecting $H_0$: „all variances are the same“ that range between 0.16 and 0.78 leading to the conclusion that the null can not be rejected for all 6 variables. Klapper (1998) shows that there is enough difference between the variances to make combining techniques based on ranks or MSE outperform the Simple Average most of the time. These methods require a difference of variances to be effective.

The correlations between the forecasters within each variable are fairly high ranging between 0.65 and 0.98, Table 1 shows the individual ranges for each variable. The distributions of the correlations are graphically displayed in Figure 1. The boxplots in the upper portion are showing a wider spread of correlations for the GDP, public consumption, and consumer prices. These variables also have several peaks, and by looking at the histograms in the lower portion of Figure 1 those different distributions possibly indicate different forecasting methods with the forecasters falling into several groups. The correlations of the other three variables are less wide spread indicating that these variables are easier to forecast.

The distributions of the correlations between variables for each forecaster are very similar. The boxplot in Figure 2 shows the majority of the correlations between 0.2 and 0.6. The histograms also look similar in shape. There are some negative correlations, the highest of which in absolute value for each forecaster is the correlation between public consumption and consumer prices. Closest to zero are the correlations between GDP and consumer prices and between private consumption and exports. The highest correlations are between GDP and imports and between exports and imports. All these relationships make perfect sense since they are in line with the underlying econometric models and therefore general economic theory.

## 2.1 Autocorrelations

The autocorrelations between variables may be of interest since there might be a dependence between one variable and a different variable in the previous time period. This dependence may be important for multivariate variable combinations. Table 2 shows the correlations between two variables and the corresponding 1-lag autocorrelations. In most cases the autocorrelations are closer to zero and only 10% of the variable pairs have an autocorrelation greater than 0.30 as opposed to 67% of the correlations. Only the autocorrelation between GDP and consumer prices is -0.63 and much higher than the corresponding correlation of 0.09. In all other cases of high autocorrelations we have higher correlations. This indicates that autocorrelations may not play an important role in this particular data set. We will therefore not consider multivariate combinations in this analysis but may consider it in the ongoing analysis process.

# Table 1 - Variances and Test of Equality

| *GDP* | |
|---|---|
| ***Variance*** | ***Scaled Variance*** |
| 2,44 | 1,97 |
| 1,75 | 1,41 |
| 1,70 | 1,37 |
| 1,49 | 1,19 |
| 1,42 | 1,15 |
| 1,42 | 1,15 |
| 1,26 | 1 |
| ***Bartlett p*** | 0.3117 |
| ***corr.*** | 0,73-0,97 |

| *Private Cons.* | |
|---|---|
| ***Variance*** | ***Scaled Variance*** |
| 2,12 | 2,04 |
| 1,72 | 1,65 |
| 1,67 | 1,61 |
| 1,24 | 1,19 |
| 1,21 | 1,16 |
| 1,12 | 1,08 |
| 1,04 | 1 |
| | 0.7519 |
| | 0,82-0,95 |

| *Public Cons.* | |
|---|---|
| ***Variance*** | ***Scaled Variance*** |
| 3,31 | 2,83 |
| 2,08 | 1,78 |
| 1,57 | 1,34 |
| 1,54 | 1,32 |
| 1,37 | 1,17 |
| 1,23 | 1,05 |
| 1,17 | 1 |
| | 0.7756 |
| | 0,65-0,95 |

| *Export* | |
|---|---|
| ***Variance*** | ***Scaled Variance*** |
| 21,5 | 1,50 |
| 19,1 | 1,34 |
| 17,4 | 1,22 |
| 15,4 | 1,08 |
| 15,3 | 1,07 |
| 15,0 | 1,05 |
| 14,3 | 1 |
| | 0.6027 |
| | 0,85-0,98 |

| *Import* | |
|---|---|
| ***Variance*** | ***Scaled Variance*** |
| 17,5 | 1,86 |
| 13,6 | 1,45 |
| 12,4 | 1,32 |
| 10,9 | 1,16 |
| 10,6 | 1,13 |
| 10,0 | 1,06 |
| 9,4 | 1 |
| | 0.1626 |
| | 0,86-0,97 |

| *Consumer Prices* | |
|---|---|
| ***Variance*** | ***Scaled Variance*** |
| 1,08 | 1,52 |
| 0,95 | 1,34 |
| 0,88 | 1,24 |
| 0,79 | 1,11 |
| 0,78 | 1,10 |
| 0,72 | 1,01 |
| 0,71 | 1 |
| | 0.8038 |
| | 0,72-0,96 |

*Bartlett p = p-value from Bartlett's Test of equal variances*
*corr. = range of correlations*

# Table 2 - Autocorrelations

| | | ***GDP*** | ***Private Cons.*** | ***Public Cons.*** | ***Export*** | ***Import*** | ***Consumer Prices*** |
|---|---|---|---|---|---|---|---|
| ***GDP*** | ***Corr*** | - | 0,46 | 0,42 | 0,73 | 0,80 | 0,09 |
| | ***1-Lag AC*** | - | 0,02 | -0,06 | -0,19 | 0,25 | -0,63 |
| ***Private Cons.*** | ***Corr*** | 0,46 | - | 0,37 | -0,04 | 0,34 | 0,56 |
| | ***1-Lag AC*** | -0,14 | - | 0,01 | 0,05 | -0,24 | 0,03 |
| ***Public Cons.*** | ***Corr*** | 0,42 | 0,37 | - | 0,25 | 0,41 | -0,11 |
| | ***1-Lag AC*** | 0,10 | 0,11 | - | -0,10 | -0,13 | -0,30 |
| ***Export*** | ***Corr*** | 0,73 | -0,04 | 0,25 | - | 0,79 | 0,53 |
| | ***1-Lag AC*** | -0,18 | 0,15 | -0,28 | - | -0,39 | -0,30 |
| ***Import*** | ***Corr*** | 0,80 | 0,34 | 0,41 | 0,79 | - | 0,26 |
| | ***1-Lag AC*** | -0,13 | -0,22 | -0,22 | 0,00 | - | 0,01 |
| ***Consumer Pr.*** | ***Corr*** | 0,09 | 0,56 | -0,11 | 0,53 | 0,26 | - |
| | ***1-Lag AC*** | -0,18 | -0,02 | -0,22 | -0,07 | -0,16 | - |

# 3. Simulation Study

Our data set of macro economic forecasts contains six variables that have a certain variance-covariance structure presented in Chapter 2. The purpose of this simulation study will be to find out which structure favors which combining technique. We will generate a dataset using a specific variance-covariance structure. In this dataset we use 10 time periods as a performance window to calculate the combining weights and apply these weights to the 12th time period to calculate the combining errors as explained in Klapper (1998). This is done step-by-step until the last forecast in the dataset is reached. We then calculate a criterion like the MSE for the combining errors of each technique.

This is repeated a certain number e.g. 100 times. Finally we determine how often a combining technique results in a lower MSE than the MSE using the Simple Average and compare the average MSE over all runs for each combining technique to the average MSE using the Simple Average. The Simple Average is used as a comparison basis because it is commonly used in literature and very popular since it is simple to calculate. Its weights add to one and are greater than zero and it is the optimal combining technique if all variances are the same, but it does not depend on the quality of past forecasts. The combining techniques to be compared to the Simple Average are:

<u>Rank based 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. This results in coefficients for each forecaster that are greater than zero and add up to one, which is the same for all other rank based methods.
- The RQua technique does the same as Rank but takes the quadrupled ranks instead of the simple ranks.
- RHist works like Rank but includes all past time points instead of the last 10.
- R0.5 averages the coefficients calculated like Rank and the coefficients of the previous time period.
- cciv and cciv3 only consider the ranks of the best or the best 3 forecasters, respectively. They are explained in detail in Russell and Adam (1987).

<u>Other combining methods:</u>

- 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 cmad technique works like cmse but uses the MAD instead of the MSE.
- 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.

## 3.1 Preparing for the Simulation

First we want to determine which criterion we are going to use to compare the forecast combination errors. To be close to our real data we will assume the forecast errors of seven forecasters to be multivariate normal distributed with variances ranging from 1 to 3 and all correlations being 0.85. We also want to find out the optimal number of repeats for our simulation which is important because computing time can get very extensive due to loops necessary for some of the combining techniques.

Table 3 shows the results and is subdivided into 4 subtables. The first and second subtable use the MSE as a comparison measure. These two tables stand for two different needs of quality for the combination of forecasts. If the question is how often the Simple Average is beaten regardless if the margin is slim or huge, the first subtable should be used. The second subtable follows the idea that the average MSE of the forecast combination errors should be as low as possible in the average. This table is less robust towards outliers since extremely bad performances have more influence on the outcome of the second subtable. Since both aspects are of interest to us, we will consider both tables to determine a rank ordering of forecast combining techniques. The third and fourth subtable are using the MAD as a comparison measure, respectively.

The first and the third subtable indicate the percentage of times the listed techniques beat the Simple Average. For better understanding we display the number of times the Simple Average is not beaten. A zero indicates that a combining technique performs 100% better than the Simple Average. The second and fourth subtable show the mean criterion for each combining technique in relation to the Simple Average's mean MSE or MAD, respectively. The percentage difference to the Simple Average's mean MSE is shown with a 5 meaning a 5% lower MSE and a -6 standing for 6% higher MSE than the Simple Average's MSE. All numbers in Table 4a and 4b are rounded numbers. In each of the four subtables we use 100, 500, 1,000, and 2,000 runs to determine the optimal number of times to repeat the simulation. The techniques are sorted in each subtable by the numbers of the 2,000-row, the first and third subtable ascending because lower means better and the other two subtables descending since the higher the percentage of the corresponding criterion, the better the combination technique.

To determine whether to use the MSE or MAD, we compare the second and fourth subtable that show the same ranking of combining techniques for 2,000 runs. Using 100 runs, there are some changes in the rank ordering of combining techniques, but using 500+ there are only two rank order changes between cciv and cmse: For 1,000 runs in the second and for 500 runs in the fourth subtable. The performance of both methods is very similar and rank ordering differences may be due to chance. In the first and third subtables the RQua, RHis, and cmad techniques perform almost equally well leading in some rank ordering changes that could be due to chance. Altogether we can conclude that MSE and MAD lead to the same results and we can therefore concentrate more on using the more popular MSE criterion used in the first and second subtable.

In the first subtable are three incidences with pairwise switches in the rank ordering. The switch at 100 runs between cmse and RQua is the most severe. The switches beween RQua, RHis, and cmad that range between 2.4 and 3.6 for 500 runs could be due to chance as discussed before since the numbers are very similar. In the second subtable there are also changes in rank ordering beween cciv and cmse for 100 and 1,000 runs. All this leads to the conclusion that we should use at least 500 runs and we should not interprete slight differences in performance as being significant for making a combining technique superior to another.

## 3.2 Simulation Results

The Simulation results are displayed in Table 4. Table 4a shows the percentage of times the Simple Average is not beaten and Table 4b shows the percentage of the mean MSE of each technique compared to the Simple Average's mean MSE. We will consider a variety of variance-covariance structures and therefore vary 3 parameters: The spread of (constant) variances, the correlation, and the length of the time series of forecasts considered. We will start with 7 forecasters to be combined and also try 2, 3, and 5 forecasters. As a second step we will make the variances to change over time. In this case the variances are considered as stochastic processes with a probability to change for each time period. There are maximum and minimum boundaries for the variances to keep them in a realistic range. This change of the variances simulates the change in the forecasting process of each forecaster that can be caused by the use of new forecasting methods, new staff or new political interests of the forecaster.

The combining techniques are sorted by the performance of the basic case in row 1 named simulation 1, where 7 forecasters forecast at 20 time points with variances 3, 2.5, 2.5, 2, 2, 1.5, 1 and correlations 0.85. There is no autocorrelation assumed to be present. We can see from Table 4b that the RQua method has the lowest average MSE followed by cciv and cmse. Table 4a shows that the cmse and cmad combination techniques based on Russell and Adam (1987) beat the Simple Average most often followed by the rank based method RQua and the more conservative rank techniques RHis and R0.5 that consider all previous time periods or respectively weight the previous weights with 0.5. As opposed to using the real data, the simple rank technique Rank performs worse than these methods because the covariance matrix is held constant. This indicates that the variances of the real data may not be constant. The New/Gr method performs significantly worse than all other methods. Table 4a shows that it does not outperform the Simple Average 62% of the time compared to 18% of the worst performing rank based combination technique. Looking at Table 4b we can see that its mean MSE is 17% higher than the Simple Average's mean MSE. The rank methods are ranging between 2% and 6% lower than the Simple Average's mean MSE. A reason for this bad performance of the New/Gr method could be that it does not constrain its weights to be greater than zero. Therefore it sometimes produces extreme positive and negative weights that come up with nonsense combined values. Considering the standard deviations of the MSEs in Table 4b we can test if the differences from the Simple Average's MSE are significant. Using a simple t-test to test $H_0$: „Two means are equal", results in significant differences for all deviances greater or equal 3% and some of the 2% deviances. Significant differences are marked with a star in Table 4b.

If we now vary the spread of variances and compare simulations 1-4, we can see no significant change in the rank ordering of the performance of the combining techniques except the New/Gr method. Generally for all methods, the higher the variance spread the higher the number of times the Simple Average is outperformed. Table 4b also shows that the higher the variance spread the closer the New/Gr methods gets to the rank based techniques. In simulation 4 with a variance spread of 7 to 1, seven methods beat the Simple Average all the time. In this scenario the New/Gr method yields the lowest average MSE with 35% less than the Simple Average's mean MSE outperforming the rank and MSE based methods by a margin of more than 10 percentage points.

In simulations 1 and 5-11 we vary the correlation. The lower the correlation the worse the peformance of all methods. The rank ordering of combining techniques changes with lowering the correlation because the decrease in performance does not happen equally fast for all methods. For a correlation of less than 0.6 the rank based methods beat the Simple Average more often than the cmse and cmad methods with the method Rank being the best for correlations of 0.2 and less in both tables. The New/Gr method performs significantly worse than all other methods when the correlations are lower. For a correlation of 0.95 its mean MSE is with 20% improvement over the Simple Average the lowest for all combining techniques but for a correlation of 0.90 it is already worse than all other techniques including the Simple Average.

Comparing simulations 1 and 12 shows that increasing the length of the time series from 20 to 50 does increase the number of times the MSE is lower than the Simple Average's MSE for all methods except the New/Gr as displayed in Table 4a. Table 4b reveals that the average MSE does not change significantly by increasing the time series length. The explanation for the good performance of rank based and MSE based combining techniques could be the ability of these methods to use more time points to „learn“ about the variance-covariance structure. The New/Gr method performs even worse because its bad performance usually results from some extreme errors caused by the problems mentioned above. The longer the time series the higher the probability of these extreme errors to happen at least once.

We now vary the number of forecasts to be combined. Looking at simulations 13-15 and 1 we can see that the New/Gr method performs better if we combine a smaller number of forecasters but even combining 2 forecasts it has only 52% of the time a lower MSE than the Simple Average as opposed to the best performing rank based method RHis with 73%. For combining less than 7 variances the RQua method has the lowest average. The cmse beats the Simple Average most often for combining less than 7 forecasts.

As already mentioned there are some motivations for the assumption that the variances may not be constant over time due to a variety of reasons. We now consider the variances to follow a stochastic process with a probability for the variances to change and upper and lower boundaries. Simulations 16-19 show four different scenarios where we vary both parameters, the probability to change and the boundaries. In all 4 simulations with nonconstant variances the RQua, cmse, and cmad techniques beat the Simple Averge most often. The RQua method has also always the lowest average MSE and the cmse and cciv share that position for lower change probabilities of 0.2 and 0.3. The method RHis that includes all time points does not perform significantly worse than the method Rank that only inludes the last 10 indicating the assumed change of variances may not be severe enough.

## 3.3 Theory vs. Simulation

We can use the Pitman-closeness theory as in Wenzel (1998) to calculate a probability that a certain combination technique is better than another. Table 5 shows two comparisons: cmse vs. Simple Average (SA) and New/Gr versus Simple Average. The New/Gr method explained in Newbold and Granger (1974) is a very popular method in literature but does not perform very well when more than two forecasters are combined due to a lack of constraining the coefficients to be greater than zero. In our simulation study in Simulation 1, the New/Gr method of combining forecasts was only 38.2% of the times better than the Simple Average as can be seen in Table 4a and the cmse method outperformed the Simple Average 97.6% of the time in the same simulation scenario. Looking at Table 5 we can see that the theoretical values obtained by the Pitman-closeness theory are 86.6% and 90.0%, respectively. A cause of this difference may be the poor estimation of the covariance matrix. But even if we assume the covariance matrix to be known during the simulation study which keeps the weights constant over time, the probabilities for the cmse and the New/Gr method beat the MSE of the Simple Average 99.5% and 95.7% of the time, respectively. As expected both these percentages are higher compared to the case with the covariance matrix being estimated. For the New/Gr combining technique the percentage is closer to the theoretical probability but for the cmse technique it is not. We tried other simulation scenarios by varying the variance-covariance structure as displayed in Table 5. In all cases the theoretical and simulated probabilities are significantly different with lower theoretic probabilities than the simulations with known covariance matrix.

# 4. Summary

The performance of forecast combining techniques highly depends on the variance-covariance structure of the forecast errors. The method RQua based on the ranks of the forecast errors and the method cmse based on the MSE perform best for most of the simulation scenarios tested. The New/Gr technique beats the Simple Average least often in all scenarios and is only superior in average MSE for big variance spreads and extremely high correlations. Lower correlations favor more conservative rank based combining methods like RHis, R0.5, and Rank. For extreme cases with correlations of 0.95, a large variance spread from 7 to 1, or long time series, most rank and MSE based methods always have a lower MSE than the Simple Average. The Simple Average's MSE is beaten at least 50% of the time except for zero correlation for all rank and MSE based methods except cciv. All rank and MSE based methods except cciv also always have a lower average MSE than the Simple Average except for zero correlation. The improvement of these methods over the Simple Average ranges from 1% to as much as 22% for RQua in the case of a high variance spread.

# Table 3 - Determination of Criteria and Number of Repeats

## Percentage of Times the MSE of Simple Average is not Beaten

| Simu- | | Techniques | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| lation | Iterations | cmse | RQua | cmad | RHis | R0.5 | Rank | cciv3 | cciv | New/Gr |
| 1 | 100 | 3,0 | 2,0 | 3,0 | 3,0 | 4,0 | 4,0 | 5,0 | 18,0 | 66,0 |
| 2 | 500 | 2,4 | 2,8 | 2,4 | 3,6 | 3,6 | 4,2 | 8,0 | 17,8 | 61,8 |
| 3 | 1000 | 1,8 | 2,5 | 2,2 | 2,3 | 3,6 | 4,7 | 8,9 | 20,6 | 64,2 |
| 4 | 2000 | 1,9 | 2,0 | 2,1 | 3,3 | 3,5 | 3,6 | 7,0 | 19,1 | 64,0 |

## Mean MSE as Percentage of Simple Average's Mean MSE

| Simu- | | Techniques | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| lation | Iterations | RQua | cciv | cmse | cciv3 | cmad | R0.5 | Rank | RHis | New/Gr |
| 1 | 100 | 6,2 | 5,6 | 5,7 | 5,3 | 2,9 | 2,0 | 2,0 | 2,0 | -20,4 |
| 2 | 500 | 6,4 | 6,0 | 5,9 | 5,4 | 2,9 | 2,0 | 2,0 | 2,0 | -17,0 |
| 3 | 1000 | 6,5 | 5,6 | 5,8 | 5,1 | 2,9 | 2,0 | 2,0 | 2,0 | -19,7 |
| 4 | 2000 | 6,4 | 5,9 | 5,9 | 5,3 | 2,9 | 2,0 | 2,0 | 2,0 | -21,1 |

## Percentage of Times the MAD of Simple Average is not Beaten

| Simu- | | Techniques | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| lation | Iterations | cmse | cmad | RQua | RHis | R0.5 | Rank | cciv3 | cciv | New/Gr |
| 1 | 100 | 3,0 | 4,0 | 5,0 | 7,0 | 8,0 | 6,0 | 16,0 | 28,0 | 66,0 |
| 2 | 500 | 4,2 | 4,8 | 5,0 | 4,8 | 6,2 | 7,6 | 11,0 | 23,6 | 62,6 |
| 3 | 1000 | 4,4 | 5,3 | 5,3 | 4,8 | 6,9 | 7,7 | 11,1 | 21,5 | 63,2 |
| 4 | 2000 | 3,0 | 3,8 | 4,1 | 5,5 | 6,5 | 7,5 | 10,9 | 21,9 | 61,5 |

## Mean MAD as Percentage of Simple Average's Mean MSE

| Simu-lation | Iterations | Techniques | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | RQua | cciv | cmse | cciv3 | cmad | R0.5 | Rank | RHis | New/Gr |
| 1 | 100 | 6,1 | 5,0 | 5,3 | 4,8 | 2,6 | 1,8 | 1,9 | 1,9 | -19,1 |
| 2 | 500 | 6,3 | 5,6 | 5,7 | 5,1 | 2,8 | 1,9 | 1,9 | 1,9 | -18,8 |
| 3 | 1000 | 6,3 | 6,0 | 5,7 | 5,2 | 2,8 | 2,0 | 2,0 | 1,9 | -18,2 |
| 4 | 2000 | 6,5 | 6,0 | 6,0 | 5,4 | 2,9 | 2,0 | 2,0 | 2,0 | -15,4 |

*All simulations use 7 forecasters with variances 3,2.5,2.5,2,2,1.5,1, a correlation of 0.85 and a total time series length of 20. First and third subtable: Percent of times a technique's MSE is larger than the SA's MSE.*

# Table 4a - Percentage of Times SA is Beaten

| Simu-lation | Length | Iterations | Variances | Corre-lation | Techniques cmse | cmad | RQua | RHis | R0.5 | Rank | cciv3 | cciv | New/Gr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 20 | 500 | 3, 2.5, 2.5, 2, 2, 1.5, 1 | 0,85 | 2 | 2 | 3 | 4 | 4 | 4 | 8 | 18 | 62 |
| 2 | 20 | 500 | 2, 1.5, 1.5, 1.5, 1, 1, 1 | 0,85 | 7 | 9 | 8 | 12 | 14 | 15 | 19 | 32 | 79 |
| 3 | 20 | 500 | 4, 3.5, 3, 2.5, 2, 1.5, 1 | 0,85 | 0 | 0 | 1 | 0 | 0 | 1 | 2 | 9 | 43 |
| 4 | 20 | 500 | 7, 6, 5, 4, 3, 2, 1 | 0,85 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 13 |
| 5 | 20 | 500 | 3, 2.5, 2.5, 2, 2, 1.5, 1 | 0,95 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 25 |
| 6 | 20 | 500 | " | 0,9 | 0 | 1 | 0 | 1 | 1 | 2 | 3 | 11 | 49 |
| 7 | 20 | 500 | " | 0,8 | 4 | 4 | 5 | 5 | 6 | 4 | 14 | 27 | 74 |
| 8 | 20 | 500 | " | 0,6 | 16 | 15 | 17 | 18 | 20 | 18 | 33 | 45 | 86 |
| 9 | 20 | 500 | " | 0,4 | 33 | 27 | 34 | 26 | 26 | 25 | 38 | 54 | 92 |
| 10 | 20 | 500 | " | 0,2 | 43 | 37 | 47 | 37 | 37 | 35 | 51 | 64 | 93 |
| 11 | 20 | 500 | " | 0 | 58 | 44 | 61 | 42 | 43 | 41 | 59 | 79 | 95 |
| 12 | 50 | 500 | " | 0,85 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 82 |
| 13 | 20 | 500 | 1.5,1 | 0,85 | 31 | 28 | 33 | 27 | 30 | 30 | 30 | 42 | 48 |
| 14 | 20 | 500 | 2,1.5,1 | 0,85 | 16 | 17 | 18 | 24 | 24 | 22 | 22 | 28 | 47 |
| 15 | 20 | 500 | 3,2,2,1.5,1 | 0,85 | 4 | 6 | 7 | 7 | 7 | 8 | 22 | 22 | 40 |
| | | | Var pCh min max | | | | | | | | | | |
| 16 | 20 | 500 | s. 1 0.2 1 4 | 0,85 | 4 | 4 | 4 | 6 | 6 | 6 | 10 | 20 | 60 |
| 17 | 20 | 500 | s. 1 0.3 1 4 | 0,85 | 4 | 4 | 4 | 7 | 6 | 6 | 10 | 21 | 62 |
| 18 | 20 | 500 | s. 1 0.4 1 4 | 0,85 | 3 | 3 | 3 | 4 | 5 | 5 | 9 | 22 | 58 |
| 19 | 20 | 500 | s. 1 0.3 0.5 5 | 0,85 | 3 | 3 | 3 | 5 | 3 | 3 | 5 | 13 | 40 |

*Rounded percent of times a technique's MSE is larger than the SA's MSE.*

# Table 4b - Percentage of SA's Mean MSE

| Simu-lation | Length | Iterations | Variances | Corre-lation | Techniques RQua | cciv | cmse | cciv3 | cmad | R0.5 | Rank | RHis | New/Gr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 20 | 500 | 3, 2.5, 2.5, 2, 2, 1.5, 1 | 0,85 | 6* | 6* | 6* | 5* | 3* | 2 | 2 | 2 | -17* |
| 2 | 20 | 500 | 2, 1.5, 1.5, 1.5, 1, 1, 1 | 0,85 | 3* | 3* | 3* | 3* | 2 | 1 | 1 | 1 | -45* |
| 3 | 20 | 500 | 4, 3.5, 3, 2.5, 2, 1.5, 1 | 0,85 | 11* | 11* | 10* | 10* | 5* | 4* | 4* | 4* | -6* |
| 4 | 20 | 500 | 7, 6, 5, 4, 3, 2, 1 | 0,85 | 22* | 18* | 20* | 15* | 10* | 7* | 7* | 6* | 35* |
| 5 | 20 | 500 | 3, 2.5, 2.5, 2, 2, 1.5, 1 | 0,95 | 7* | 11* | 6* | 8* | 3* | 2* | 2* | 2* | 20* |
| 6 | 20 | 500 | " | 0,9 | 7* | 8* | 6* | 7* | 3* | 2* | 2* | 2* | -4* |
| 7 | 20 | 500 | " | 0,8 | 6* | 4* | 6* | 5* | 3* | 2 | 2 | 2 | -35* |
| 8 | 20 | 500 | " | 0,6 | 6* | 1 | 6* | 3* | 3* | 2 | 2 | 2 | -56* |
| 9 | 20 | 500 | " | 0,4 | 4* | -2 | 4* | 2 | 3* | 2 | 2 | 2 | -66* |
| 10 | 20 | 500 | " | 0,2 | 1 | -6* | 2 | 0 | 2 | 2 | 2 | 2 | -76* |
| 11 | 20 | 500 | " | 0 | -7* | -22* | -4* | -4* | 1 | 1 | 2 | 2 | -81* |
| 12 | 50 | 500 | " | 0,85 | 7* | 6* | 6* | 5* | 3* | 2* | 2* | 2* | -23* |
| 13 | 20 | 500 | 1.5,1 | 0,85 | 2 | 2* | 2 | 1 | 1 | 1 | 1 | 1 | 0 |
| 14 | 20 | 500 | 2,1.5,1 | 0,85 | 3* | 3* | 3* | 1 | 2 | 1 | 1 | 1 | 2 |
| 15 | 20 | 500 | 3,2,2,1.5,1 | 0,85 | 6* | 6* | 6* | 6* | 3* | 2* | 2* | 2* | 2 |
| | | | Var pCh min max | | | | | | | | | | |
| 16 | 20 | 500 | s. 1 0.2 1 4 | 0,85 | 7* | 7* | 7* | 6* | 3* | 2* | 2* | 2 | -19* |
| 17 | 20 | 500 | s. 1 0.3 1 4 | 0,85 | 7* | 7* | 7* | 6* | 4* | 2* | 2* | 2* | -19* |
| 18 | 20 | 500 | s. 1 0.4 1 4 | 0,85 | 8* | 7* | 7* | 7* | 4* | 3* | 3* | 2* | -18* |
| 19 | 20 | 500 | s. 1 0.3 0.5 5 | 0,85 | 14* | 13* | 13* | 11* | 7* | 5* | 4* | 4* | 1 |

*Rounded percent the MSE of a technique is better than SA's MSE, negative numbers mean worse than SA's MSE. * indicates significant difference to Simple Average's mean MSE.*

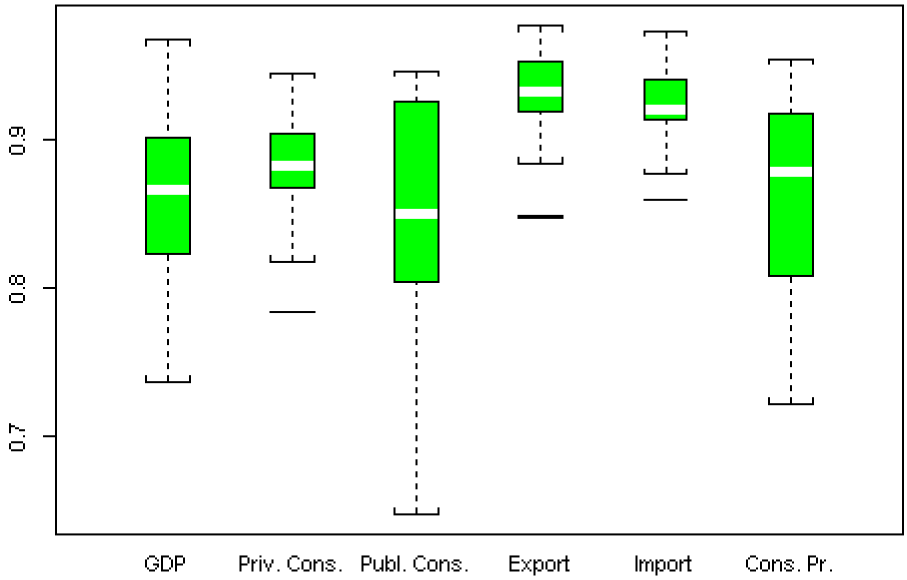
*Legend: Var = Variances, pCh = probability for variance to change, min = minimum variance, max = maximum variance.*

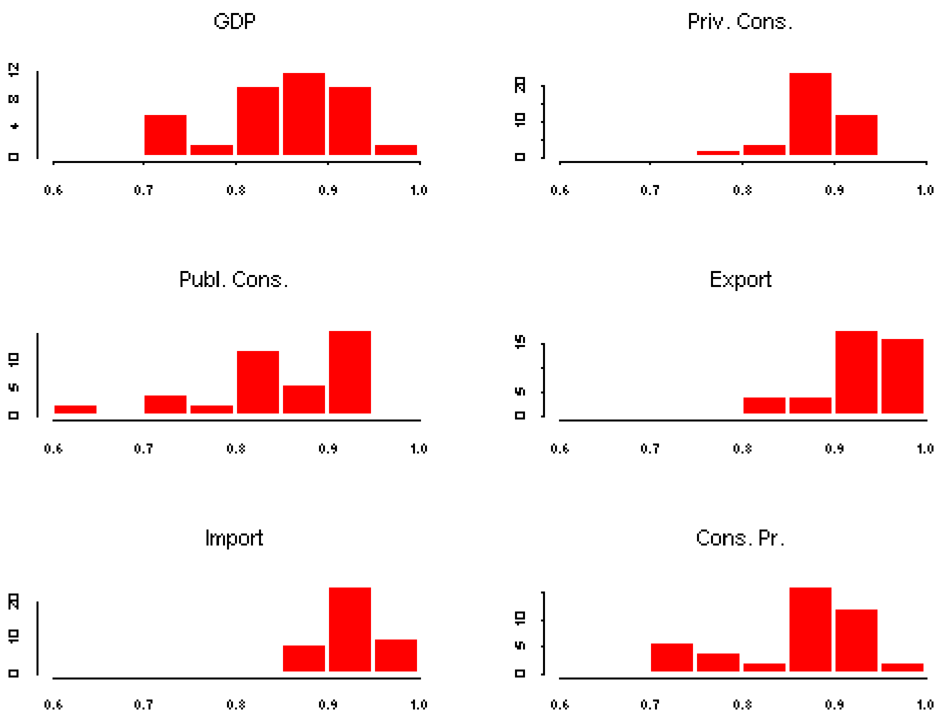
# Table 5 - Pitman Closeness Probabilities

| ***Variance structure*** | ***Corr*** | ***MSE vs. SA*** | | | ***New/Gr vs. SA*** | | |
|---|---|---|---|---|---|---|---|
| | | ***theor.*** | ***simul.*** | ***si/kn*** | ***theor.*** | ***simul.*** | ***si/kn*** |
| 3,2.5,2.5,2,2,1.5,1 | 0,85 | 0,900 | 0,976 | 0,995 | 0,866 | 0,382 | 0,957 |
| 4,3.5,3,2.5,2,1.5,1 | 0,85 | 0,930 | 1,000 | 0,999 | 0,902 | 0,570 | 0,984 |
| 3,2.5,2.5,2,2,1.5,1 | 0,95 | 0,943 | 1,000 | 1,000 | 0,910 | 0,750 | 0,999 |
| " | 0,9 | 0,920 | 0,996 | 1,000 | 0,883 | 0,512 | 0,979 |
| " | 0,8 | 0,884 | 0,958 | 0,993 | 0,852 | 0,260 | 0,919 |
| " | 0,6 | 0,830 | 0,836 | 0,970 | 0,807 | 0,144 | 0,890 |
| " | 0,4 | 0,783 | 0,674 | 0,870 | 0,766 | 0,082 | 0,760 |
| " | 0,2 | 0,721 | 0,568 | 0,760 | 0,709 | 0,066 | 0,680 |
| " | 0 | 0,556 | 0,424 | 0,740 | 0,556 | 0,050 | 0,740 |
| 3,1 | 0,85 | 0,742 | 0,690 | 0,990 | 0,660 | 0,520 | 0,940 |
| 3,2,1 | 0,85 | 0,747 | 0,840 | 0,996 | 0,659 | 0,530 | 0,936 |
| 3,2,2,1.5,1 | 0,85 | 0,750 | 0,960 | 0,991 | 0,657 | 0,600 | 0,928 |

*si/kn = simulation, known covariance matrix for weight determination.*

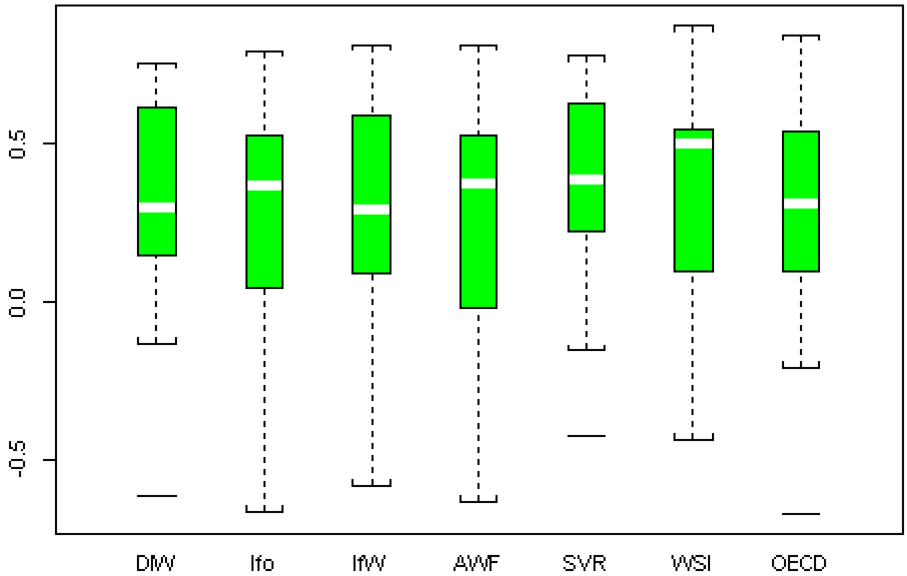
# Figure 1 - Distribution of the Correlations

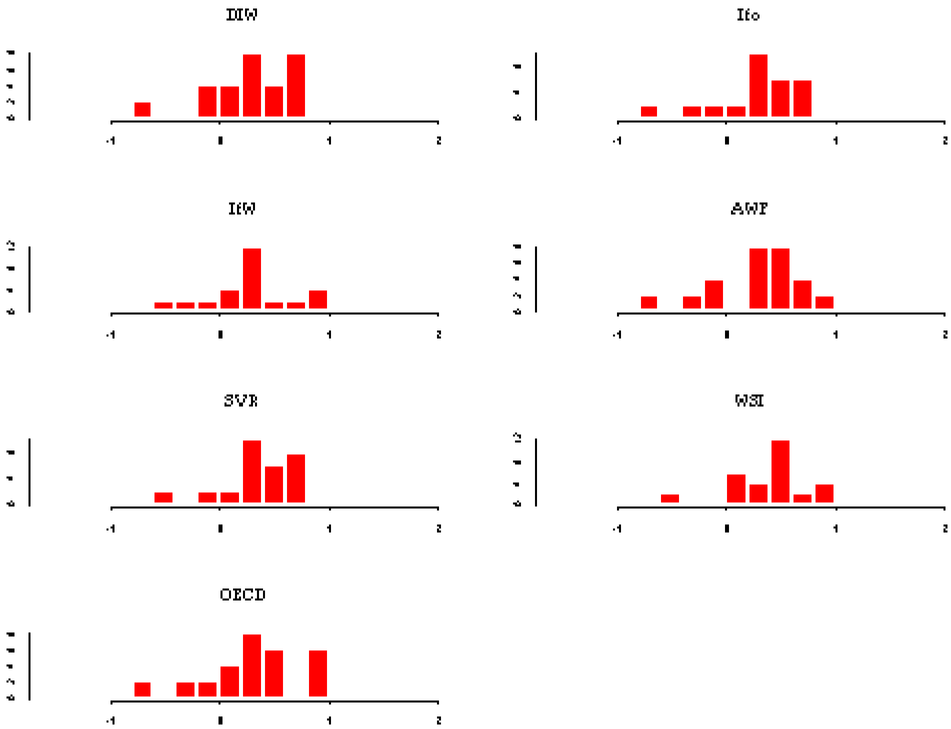



*The distribution of the correlations between the 7 forecasters for the 6 variables.*

# Figure 2 - Between-Variable Correlations




*The distribution of the between-variable correlations for each of the 7 forecaster..*

# Bibliography


- Klapper, Matthias (1998): Combining German Macro Economic Forecasts using Rank-Based Techniques, Technical Report 19/1998, 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 as a Measure to Evaluate the Quality of Forecasts, Technical Report 15/1998, SFB 475, University of Dortmund


# Appendix A

## Bartlett's Test

Test of $H_0$: „The variances are unequal."

$$X^2 = \frac{2.303}{C}\left( n \log(\bar{S})^2 - \sum n_i \log(S_i^2) \right)$$

with $C = 1 + \frac{1}{3(k-1)}\left( \sum \frac{1}{n_i} - \frac{1}{n} \right)$,

and with the $S_i$ being the individual variances, S the pooled variance and n the the sum of all $n_i$.