The shrinkage approach in the combination of forecasts

Thomas Wenzel

Department of Statistics, University of Dortmund, Vogelpothsweg 87

D-44227 Dortmund, Germany

e-mail: wenzel@statistik.uni-dortmund.de

Abstract: An unbiased point estimator T for an unknown parameter θ can be improved in the sense of the Mean Squared Error (MSE) by $T_\lambda = \lambda T$ for suitable factors λ. Here, we want to discuss this approach in the context of combination of forecasts. We consider the shrinkage technique for unbiased univariate and multivariate forecast combinations. In the univariate case our aim is to reduce the MSE. In the multivariate case we want to improve unbiased forecast combinations in the sense of the Scalar Mean Squared Error (SMSE) or the Matrix Mean Squared Error (MMSE).

Keywords: shrinkage, combination of forecast, mean squared error.

Acknowledgement: Financial support of the Deutsche Forschungsgemeinschaft (SFB 475, “Reduction of complexity in multivariate data structures“) is gratefully acknowledged. M. Hibon kindly provided us with a copy of the data from the M-competition.

AMS 1991 Subject classification: 62F10

1. Introduction

The most popular forecasting evaluation criterion is the Mean Squared Error (MSE). Thus, the combination of forecasts is commonly based on the MSE. In general we assume that the individual forecasts are unbiased which means that there are no systematical errors. Based on this, forecast combinations are often restricted to be unbiased, resulting in the restriction that the weights sum up to 1 (see e.g. Bates and Granger (1969) or Dickinson (1973)). Giving up the unbiasedness restriction can lead to forecast combinations with smaller MSE. In this case the optimal combination weights depend on the second moment of the variable to be forecasted. This makes it difficult to estimate the weights in many applications. Another approach is to shrink unbiased forecast combinations. This also results in a dependence of the variable we are forecasting but it is possible to calculate the size of estimation errors still leading to an improvement. Further, we can see that optimal shrinking of the optimal unbiased combination is equivalent to the MSE-optimal technique. Using some data from the M-competition (Makridakis et al. (1982)) we analyse the quality of different unbiased combination techniques and their optimal shrunken versions.

Furthermore, we want to discuss the multivariate case. Here, the comparison of forecasting techniques is usually based on the Matrix Mean Squared Error (MMSE) or on its trace, the Scalar Mean Squared Error (SMSE). We consider two different shrinkage approaches. The first is based on a shrinkage scalar λ and the second on a shrinkage matrix Γ. We calculate optimal combinations in the sense of the SMSE and the MMSE. For a better illustration we perform a simulation study for the multivariate case.

2. The univariate case

We consider the following situation **(S1)**: Let $F_{i,T+L}$, $i = 1,...,n$ be unbiased forecasts for Y_{T+L} at time T, where $L \in \mathbb{N}$ denotes the forecast horizon. Thus, $E(u_{i,T+L}) = 0$, where $u_{i,T+L} := Y_{T+L} - F_{i,T+L}$, $i = 1,...,n$. Furthermore, we assume that $\mathrm{Cov}(Y_{T+L}, u_{i,T+L}) = 0$, $i = 1,...,n$, and $\mathrm{Cov}(\mathbf{u}_{T+L}) =: \Sigma$ is p.d., where $\mathbf{u}_{T+L} := (u_{1,T+L},...,u_{n,T+L})'$.

Assuming that $E(Y_{T+L}) \neq 0$, it is well-known, that the MSE-optimal unbiased forecast combination of the n individual forecasts is given by

$$F_{\mathbf{c}_{opt,unb},T+L} := \mathbf{c}_{opt,unb}{}' \mathbf{F}_{T+L} \quad ,$$

where $\mathbf{c}_{opt,unb} := (\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}\Sigma^{-1}\mathbf{1}$, (1) and $\mathbf{1}$ denotes the $n\times 1$ vector of 1's, $\mathbf{F}_{T+L} := \left(F_{1,T+L},\ldots,F_{n,T+L}\right)'$. Furthermore we have $\mathrm{MSE}\left(Y_{T+L}, F_{\mathbf{c}_{opt,unb},T+L}\right) = \left(\mathbf{1}'\Sigma^{-1}\mathbf{1}\right)^{-1}$.

Now we calculate the MSE-optimal (biased) forecast combination.

<u>*Theorem 1*</u>: Considering the situation **(S1)**, the MSE-optimal forecast combination is given by $F_{\mathbf{c}_{opt,b},T+L} := \mathbf{c}_{opt,b}{}'\mathbf{F}_{T+L}$, where $\mathbf{c}_{opt,b} = \left(E\left(Y_{T+L}^2\right)\mathbf{1}'\Sigma^{-1}\mathbf{1}+1\right)^{-1} E\left(Y_{T+L}^2\right)\Sigma^{-1}\mathbf{1}$.

Proof: Let $F_{\mathbf{c},T+L} := \mathbf{c}'\mathbf{F}_{T+L}$, where $\mathbf{c} := (c_1,\ldots,c_n)'$, $\mathbf{c}'\mathbf{1} = 1 + d, d \in \mathbb{R}$ and

$$u_{\mathbf{c},T+L} := Y_{T+L} - F_{\mathbf{c},T+L} = (1+d)Y_{T+L} - \mathbf{c}'\mathbf{F}_{T+L} - dY_{T+L}$$
$$= \mathbf{c}'\mathbf{u}_{T+L} - dY_{T+L} \ .$$

Thus,

$$E\left(u_{\mathbf{c},T+L}\right) = \mathbf{c}'E\left(\mathbf{u}_{T+L}\right) - dE\left(Y_{T+L}\right) = \left(1-\mathbf{c}'\mathbf{1}\right)E\left(Y_{T+L}\right)$$

and

$$\mathrm{Var}\left(u_{\mathbf{c},T+L}\right) = \mathbf{c}'\Sigma\mathbf{c} + d^2\mathrm{Var}\left(Y_{T+L}\right) - 2\sum_{i=1}^{n} c_i d\,\mathrm{Cov}\left(u_{i,T+L}, Y_{T+L}\right)$$
$$= \mathbf{c}'\Sigma\mathbf{c} + \left(\mathbf{c}'\mathbf{1}-1\right)^2\mathrm{Var}\left(Y_{T+L}\right) \ .$$

Now we can calculate the MSE.

$$\mathrm{MSE}\left(Y_{T+L}, F_{\mathbf{c},T+L}\right) = \mathrm{Var}\left(u_{\mathbf{c},T+L}\right) + \left(E\left(u_{\mathbf{c},T+L}\right)\right)^2$$
$$= \mathbf{c}'\Sigma\mathbf{c} + \left(\mathbf{c}'\mathbf{1}-1\right)^2\left(E\left(Y_{T+L}^2\right) - E\left(Y_{T+L}\right)^2\right) + \left(\mathbf{c}'\mathbf{1}-1\right)^2 E\left(Y_{T+L}\right)^2$$
$$= \mathbf{c}'\Sigma\mathbf{c} + \left(\mathbf{c}'\mathbf{1}-1\right)^2 E\left(Y_{T+L}^2\right)$$

Since we want to minimize the MSE, we consider

$$\frac{\partial\,\mathrm{MSE}\left(Y_{T+L}, F_{\mathbf{c},T+L}\right)}{\partial\mathbf{c}} = 2\cdot\mathbf{c}'\Sigma + 2\cdot\mathbf{c}'\mathbf{1}\mathbf{1}'E\left(Y_{T+L}^2\right) - 2\cdot\mathbf{1}'E\left(Y_{T+L}^2\right) \overset{!}{=} \mathbf{0}_{n\times 1}{}'$$
$$\Leftrightarrow 2\cdot\mathbf{c}'\left(\Sigma + \mathbf{1}\mathbf{1}'E\left(Y_{T+L}^2\right)\right) = 2\cdot\mathbf{1}'E\left(Y_{T+L}^2\right)$$

and thus,

$$\mathbf{c}_{opt,b} = E\left(Y_{T+L}^2\right)\left(\Sigma + \mathbf{1}\mathbf{1}'E\left(Y_{T+L}^2\right)\right)^{-1}\mathbf{1} \ . \qquad (2)$$

Since $\dfrac{\partial^2\mathrm{MSE}\left(Y_{T+L}, F_{\mathbf{c},T+L}\right)}{\partial^2\mathbf{c}} = \Sigma + \mathbf{1}\mathbf{1}'E\left(Y_{T+L}^2\right)$ is p.d. follows that $\mathbf{c}_{opt,b}$ is the minimizing vector. Consulting Horn and Johnson (1985, p. 19) we get

$\left(\Sigma+\mathbf{11}'\mathrm{E}\left(\mathrm{Y}_{\mathrm{T+L}}^{2}\right)\right)^{-1}=\Sigma^{-1}-\left(\frac{1}{\mathrm{E}\left(\mathrm{Y}_{\mathrm{T+L}}^{2}\right)}+\mathbf{1}'\Sigma^{-1}\mathbf{1}\right)^{-1}\Sigma^{-1}\mathbf{11}'\Sigma^{-1}$ and by some easy calculations we get

$$\mathbf{c}_{\mathrm{opt,b}}=\left(\mathrm{E}\left(\mathrm{Y}_{\mathrm{T+L}}^{2}\right)\mathbf{1}'\Sigma^{-1}\mathbf{1}+1\right)^{-1}\mathrm{E}\left(\mathrm{Y}_{\mathrm{T+L}}^{2}\right)\Sigma^{-1}\mathbf{1}\quad. \tag{3}$$

Furthermore, $\mathrm{MSE}\left(\mathrm{Y}_{\mathrm{T+L}},\mathrm{F}_{\mathbf{c}_{\mathrm{opt,b}},\mathrm{T+L}}\right)=\left(\mathrm{E}\left(\mathrm{Y}_{\mathrm{T+L}}^{2}\right)\mathbf{1}'\Sigma^{-1}\mathbf{1}+1\right)^{-1}\mathrm{E}\left(\mathrm{Y}_{\mathrm{T+L}}^{2}\right)$. If $\mathrm{E}\left(\mathrm{Y}_{\mathrm{T+L}}\right)=0$ the forecast combination $\mathrm{F}_{\mathbf{c}_{\mathrm{opt,b}},\mathrm{T+L}}$ is also unbiased.

We can see that the optimal weights depend on the second moment of the variable $\mathrm{Y}_{\mathrm{T+L}}$. In practice it could be difficult to estimate this, especially when the second moment of the variable to be forecasted is not constant.

In the following we discuss the shrinkage technique for an improvement of unbiased forecast combinations. We analyse again the MSE-optimal combination in this context.

<u>*Theorem 2*</u>: Consider a forecast combination $\mathrm{F}_{\mathbf{c},\mathrm{T+L}}=\mathbf{c}'\mathbf{F}_{\mathrm{T+L}}$ in situation **(S1)**, where $\mathbf{c}'\mathbf{1}=1$ and further $\mathrm{F}_{\lambda\mathbf{c},\mathrm{T+L}}:=\lambda\mathrm{F}_{\mathbf{c},\mathrm{T+L}}$, $\lambda\in\mathbb{R}$, where $\lambda\mathbf{c}'\mathbf{1}=\lambda=1+\mathrm{k}$, $\mathrm{k}\in\mathbb{R}$. Then the MSE-minimizing λ is given by $\lambda_{\mathbf{c},\mathrm{opt}}:=\frac{\mathrm{E}\left(\mathrm{Y}_{\mathrm{T+L}}^{2}\right)}{\mathrm{E}\left(\mathrm{Y}_{\mathrm{T+L}}^{2}\right)+\mathbf{c}'\Sigma\mathbf{c}}$. The forecast combinations $\mathrm{F}_{\lambda\mathbf{c},\mathrm{T+L}}$, where $\lambda\in\left(\frac{\mathrm{E}\left(\mathrm{Y}_{\mathrm{T+L}}^{2}\right)-\mathbf{c}'\Sigma\mathbf{c}}{\mathrm{E}\left(\mathrm{Y}_{\mathrm{T+L}}^{2}\right)+\mathbf{c}'\Sigma\mathbf{c}},1\right)$, have a smaller MSE than $\mathrm{F}_{\mathbf{c},\mathrm{T+L}}$.

Proof: We get

$$\begin{aligned}\mathrm{u}_{\lambda\mathbf{c},\mathrm{T+L}}&:=\mathrm{Y}_{\mathrm{T+L}}-\mathrm{F}_{\lambda\mathbf{c},\mathrm{T+L}}=(1+\mathrm{k})\mathrm{Y}_{\mathrm{T+L}}-\lambda\mathbf{c}'\mathbf{F}_{\mathrm{T+L}}-\mathrm{k}\mathrm{Y}_{\mathrm{T+L}}\\&=\lambda\mathbf{c}'\mathbf{u}_{\mathrm{T+L}}-(\lambda-1)\mathrm{Y}_{\mathrm{T+L}}\quad.\end{aligned}$$

The mean and the variance of $\mathrm{u}_{\lambda\mathbf{c},\mathrm{T+L}}$ are given by

$$\mathrm{E}\left(\mathrm{u}_{\lambda\mathbf{c},\mathrm{T+L}}\right)=-(\lambda-1)\mathrm{E}\left(\mathrm{Y}_{\mathrm{T+L}}\right)\quad\text{and}$$

$$\mathrm{Var}\left(\mathrm{u}_{\lambda\mathbf{c},\mathrm{T+L}}\right)=\lambda^{2}\mathbf{c}'\Sigma\mathbf{c}+(\lambda-1)^{2}\mathrm{Var}\left(\mathrm{Y}_{\mathrm{T+L}}\right)\quad.$$

From this we can calculate the MSE of the forecast combination $\mathrm{F}_{\lambda\mathbf{c},\mathrm{T+L}}$.

$$\begin{aligned}\mathrm{MSE}\left(\mathrm{Y}_{\mathrm{T+L}},\mathrm{F}_{\lambda\mathbf{c},\mathrm{T+L}}\right)&=\mathrm{Var}\left(\mathrm{u}_{\lambda\mathbf{c},\mathrm{T+L}}\right)+\left(\mathrm{E}\left(\mathrm{u}_{\lambda\mathbf{c},\mathrm{T+L}}\right)\right)^{2}\\&=\lambda^{2}\mathbf{c}'\Sigma\mathbf{c}+(\lambda-1)^{2}\mathrm{E}\left(\mathrm{Y}_{\mathrm{T+L}}^{2}\right)\end{aligned}$$

Now the question arises, for which λ we get $\mathrm{MSE}\left(Y_{T+L}, F_{\lambda\mathbf{c},T+L}\right) < \mathrm{MSE}\left(Y_{T+L,} F_{\mathbf{c},T+L}\right) = \mathbf{c}'\Sigma\mathbf{c}$, that is $\lambda^2\left(\mathbf{c}'\Sigma\,\mathbf{c} + E\left(Y_{T+L}^2\right)\right) - 2\lambda E\left(Y_{T+L}^2\right) + \left(E\left(Y_{T+L}^2\right) - \mathbf{c}'\Sigma\,\mathbf{c}\right) < 0$. By some easy caculations we get the values of λ, where the left side of the former inequality is 0. Thus, the improvement region (interval) for λ with respect to the forecast combination $F_{\mathbf{c},T+L}$ is given by:

$$\mathrm{IR}_{F_{\mathbf{c},T+L}} := \left(\frac{E\left(Y_{T+L}^2\right) - \mathbf{c}'\Sigma\,\mathbf{c}}{E\left(Y_{T+L}^2\right) + \mathbf{c}'\Sigma\,\mathbf{c}}, 1 \right). \qquad (4)$$

We are further interested in $\lambda_{\mathbf{c},\mathrm{opt}}$, which minimizes the MSE. We calculate

$$\frac{\partial\, \mathrm{MSE}\left(Y_{T+L,}, F_{\lambda\mathbf{c},T+L}\right)}{\partial\,\lambda} = 2\lambda\left(\mathbf{c}'\Sigma\,\mathbf{c} + E\left(Y_{T+L}^2\right)\right) - 2E\left(Y_{T+L}^2\right) \overset{!}{=} 0$$

$$\Leftrightarrow \lambda_{\mathbf{c},\mathrm{opt}} = \frac{E\left(Y_{T+L}^2\right)}{E\left(Y_{T+L}^2\right) + \mathbf{c}'\Sigma\,\mathbf{c}} \quad ,$$

and since $\dfrac{\partial^2\, \mathrm{MSE}\left(Y_{T+L,}, F_{\lambda\mathbf{c},T+L}\right)}{\partial^2\,\lambda} > 0$, this is the minimizing λ.

It is obvious that only for $\lambda < 1$ an improvement is possible. Again, we can see that the improvement region depends on the variable to be forecasted. Looking at (4), the scalar $\lambda_{\mathbf{c},\mathrm{opt}}$ is the midpoint of the corresponding improvement region. The optimal weights in the shrunken forecast combination are given by

$$\mathbf{g}_{\mathbf{c},\mathrm{opt}} := \lambda_{\mathbf{c},\mathrm{opt}}\mathbf{c} = \left(E\left(Y_{T+L}^2\right) + \mathbf{c}'\Sigma\,\mathbf{c}\right)^{-1} E\left(Y_{T+L}^2\right)\mathbf{c}.$$

Considering $\mathbf{c} := \mathbf{c}_{\mathrm{opt,unb}} = \left(\mathbf{1}'\Sigma^{-1}\mathbf{1}\right)^{-1}\Sigma^{-1}\mathbf{1}$ as in (1), we get

$$\mathbf{g}_{\mathbf{c}_{\mathrm{opt,unb}},\mathrm{opt}} = \left(\left(E\left(Y_{T+L}^2\right) + \frac{\mathbf{1}'\Sigma^{-1}\Sigma\Sigma^{-1}\mathbf{1}}{\left(\mathbf{1}'\Sigma^{-1}\mathbf{1}\right)^2} \right) \mathbf{1}'\Sigma^{-1}\mathbf{1} \right)^{-1} E\left(Y_{T+L}^2\right)\Sigma^{-1}\mathbf{1} = \left(E\left(Y_{T+L}^2\right)\mathbf{1}'\Sigma^{-1}\mathbf{1} + 1\right)^{-1} E\left(Y_{T+L}^2\right)\Sigma^{-1}\mathbf{1}$$

$$= \mathbf{c}_{\mathrm{opt,b}} \qquad \text{(see (3))}.$$

The weights of the MSE-optimal combination and the weights of the MSE-optimal unbiased combination differ only by the factor $\lambda_{\mathbf{c}_{\mathrm{opt,unb}},\mathrm{opt}} := \left(E\left(Y_{T+L}^2\right)\mathbf{1}'\Sigma^{-1}\mathbf{1} + 1\right)^{-1} E\left(Y_{T+L}^2\right)\mathbf{1}'\Sigma^{-1}\mathbf{1}$.

The MSE of a forecast combination $F_{\lambda_{\mathbf{c},\mathrm{opt}}\mathbf{c},T+L}$ is given by

$$\mathrm{MSE}\left(Y_{T+L}, F_{\lambda_{\mathbf{c},\mathrm{opt}}\mathbf{c},T+L}\right) = \left(E\left(Y_{T+L}^2\right) + \mathbf{c}'\Sigma\,\mathbf{c}\right)^{-1} E\left(Y_{T+L}^2\right)\mathbf{c}'\Sigma\,\mathbf{c}$$

which is a strictly monotone increasing function of $\mathbf{c}'\Sigma\mathbf{c}$, and therefore a forecast combination $F_{\lambda_{\mathbf{c},opt}\mathbf{c},T+L}$ is better than $F_{\lambda_{\mathbf{z},opt}\mathbf{z},T+L}$ iff $F_{\mathbf{c},T+L}$ is better than $F_{\mathbf{z},T+L}$.

As mentioned above, in practice $E\left(Y_{T+L}^2\right)$ is unknown and must be estimated. It is important to know how large the estimation errors could be resulting still in an improvement of the given unbiased forecast combination. A realised non-negative estimator $\hat{E}\left(Y_{T+L}^2\right)$ always leads to $\hat{\lambda}_{\mathbf{c},opt} < 1$. Thus, we have to check when $\hat{\lambda}_{\mathbf{c},opt} > \frac{E\left(Y_{T+L}^2\right) - \mathbf{c}'\Sigma\mathbf{c}}{E\left(Y_{T+L}^2\right) + \mathbf{c}'\Sigma\mathbf{c}}$, the lower bound of the improvement region $IR_{F_{\mathbf{c},T+L}}$. Assuming Σ is known, for a realised non-negative $\hat{E}\left(Y_{T+L}^2\right)$ we get:

$$\hat{\lambda}_{\mathbf{c},opt} = \frac{\hat{E}\left(Y_{T+L}^2\right)}{\hat{E}\left(Y_{T+L}^2\right) + \mathbf{c}'\Sigma\mathbf{c}} > \frac{E\left(Y_{T+L}^2\right) - \mathbf{c}'\Sigma\mathbf{c}}{E\left(Y_{T+L}^2\right) + \mathbf{c}'\Sigma\mathbf{c}}$$

$$\Leftrightarrow \hat{E}\left(Y_{T+L}^2\right)\left(E\left(Y_{T+L}^2\right) + \mathbf{c}'\Sigma\mathbf{c}\right) > \left(E\left(Y_{T+L}^2\right) - \mathbf{c}'\Sigma\mathbf{c}\right)\left(\hat{E}\left(Y_{T+L}^2\right) + \mathbf{c}'\Sigma\mathbf{c}\right)$$

$$\Leftrightarrow \hat{E}\left(Y_{T+L}^2\right) > \frac{E\left(Y_{T+L}^2\right)}{2} - \frac{\mathbf{c}'\Sigma\mathbf{c}}{2} \quad . \tag{5a}$$

Hence, an underestimation of 50% of $E\left(Y_{T+L}^2\right)$ still leads to an improvement of the forecast combination $F_{\mathbf{c},T+L}$. Furthermore, if $E\left(Y_{T+L}^2\right) \le \mathbf{c}'\Sigma\mathbf{c}$, a positive $\hat{E}\left(Y_{T+L}^2\right)$ results always in a better forecast. In general Σ is also unknown and has to be estimated. This results in $\tilde{\lambda}_{\mathbf{c},opt} := \frac{\hat{E}\left(Y_{T+L}^2\right)}{\hat{E}\left(Y_{T+L}^2\right) + \mathbf{c}'\hat{\Sigma}\mathbf{c}}$, where $\hat{\Sigma}$ is a p.d. estimator of Σ. For given realised estimators this improves the unbiased forecast combination $F_{\mathbf{c},T+L}$ if

$$\tilde{\lambda}_{\mathbf{c},opt} = \frac{\hat{E}\left(Y_{T+L}^2\right)}{\hat{E}\left(Y_{T+L}^2\right) + \mathbf{c}'\hat{\Sigma}\mathbf{c}} > \frac{E\left(Y_{T+L}^2\right) - \mathbf{c}'\Sigma\mathbf{c}}{E\left(Y_{T+L}^2\right) + \mathbf{c}'\Sigma\mathbf{c}}$$

$$\Leftrightarrow \hat{E}\left(Y_{T+L}^2\right) > \frac{E\left(Y_{T+L}^2\right)\mathbf{c}'\hat{\Sigma}\mathbf{c}}{2\mathbf{c}'\Sigma\mathbf{c}} - \frac{\mathbf{c}'\hat{\Sigma}\mathbf{c}}{2} \quad . \tag{5b}$$

If $E\left(Y_{T+L}^2\right) > \mathbf{c}'\Sigma\mathbf{c}$ the right side of (5b) is a strictly monotone increasing function of $\mathbf{c}'\hat{\Sigma}\mathbf{c}$. In this case a larger estimation error of $\mathbf{c}'\Sigma\mathbf{c}$ leads to the necessity of a larger $\hat{E}\left(Y_{T+L}^2\right)$. In the case where $E\left(Y_{T+L}^2\right) \le \mathbf{c}'\Sigma\mathbf{c}$ the right side of (5b) is non-positive. Thus, a positive $\hat{E}\left(Y_{T+L}^2\right)$ leads always to an improvement. Again we wish to remark that the reduction of $MSE\left(Y_{T+L}, F_{\mathbf{c},T+L}\right)$, given by the inequalities (5a) and (5b), holds for realised estimators of the unknown parameters.

It is also possible to shrink at first the individual forecasts, which is $F_{\lambda_i,T+L} := \lambda_i F_{i,T+L}$, $\lambda_i \in \mathbb{R}$, $i = 1,...,n$. Starting from a forecast combination $F_{\mathbf{b},T+L} := \sum_{i=1}^{n} c_i \lambda_i F_{i,T+L} = \sum_{i=1}^{n} b_i F_{i,T+L}$, where $b_i := c_i \lambda_i$, $i = 1,...,n$ and minimizing the MSE obviously leads to the weights given in (3).

In the next section we analyse several unbiased forecast combinations and their shrunken versions presented above with data from the M-competition (Makridakis et al. (1982)). There we have to deal with the problem of an unknown covariance matrix Σ which makes it more difficult to get an estimator $\hat{E}(Y_{T+L}^2)$ leading to a combination with smaller MSE.

3. Application for the univariate case

We use the monthly data of the M-competition as decribed in Klapper and Wenzel (1998). The time series are of length 18. The calculation of the first combination weights is based on the first ten data points. Thus, 8 data points are left for the comparison of the methods. In each step we calculate new weights on the basis of the 10 most recent data points. We consider only four individual forecasts, that are different smoothing techniques (AEP, Bays, Holt, Quadr). For a detailed decription of these methods see Makridakis et al. (1982). We assume that the individual forecasts are unbiased. We compare the RMSE of the different forecasting methods with the RMSE of the simple average of the individual forecasts. We have to remark that we eliminated five time series because of singular $\hat{\Sigma}$. Hence, 612 time series are left for our analysis. We consider in each step $\hat{\Sigma} := (\hat{\sigma}_{ij})_{i,j=1,...,n}$, $\hat{\sigma}_{ij} := \frac{1}{10}\mathbf{u}_i{}'\mathbf{u}_j$, $i = 1,...,n$, and $\mathbf{u}_i$ denotes the vector of the most recent 10 forecast errors of the i-th individual forecast. For the estimation of $E(Y_{T+L}^2)$ we use the mean of the squared most recent 10 data points of the variable to be forecasted. We analyse the following unbiased combination techniques:

Method 1 (M1): MSE-optimal unbiased forecast combination (see (1)).
Method 2 (M2): Method 1 with the further restriction, that the weights are non-negative.
Method 3 (SA): The simple average of the individual forecasts.

Methods No. 4, 5 and 6 (denoted by S-M1, S-M2 and S-SA) are the optimal shrunken versions of methods No. 1, 2 and 3. Thus, method No. 4 is the MSE-optimal given by the weights in (3). Together with the individual forecasts and their shrunken versions (S-AEP, S-

Bays, S-Holt, S-Quadr) we focus on 14 different techniques. The results of the study are given in the following table.

Table 1: Results for the study of the 612 time series

	# better than SA	# better than best individual	# best method	mean of relative RMSEs	median of relative RMSEs
AEP	282	-	21	1.209	1.038
Bays	310	-	47	1.164	0.996
Holt	319	-	43	1.116	0.978
Quadr.	142	-	25	1.920	1.355
S-AEP	312	168	25	1.149	0.965
S-Bays	349	215	29	1.041	0.944
S-Holt	364	227	45	1.025	0.920
S-Quadr	191	127	24	1.639	1.223
M1	444	286	124	0.746	0.665
M2	481	139	34	0.762	0.777
SA	-	89	29	1.000	1.000
S-M1	446	285	105	0.745	0.666
S-M2	483	249	36	0.748	0.738
S-SA	330	148	25	0.936	0.991

The relative RMSE is given by the RMSE of a special method divided by the RMSE of the simple average combination.

At first we can say that all combination techniques are doing well. They often outperform all individual forecasts. Looking at the number of times the certain methods are best, we can see that method M1 and its shrunken version (S-M1) are the forecasts of highest quality. This is also underlined by the mean and median of the relative RMSEs. Method SA is outperformed by shrinking. Furthermore, the simple average combination is in this study the combination method of lowest quality. Each other combination method outperforms the simple average in over 50% of the given 612 time series.

Finally, the method S-M1 outperforms the method M1 in 299 cases, the method S-M2 is in 304 cases better than M2 and S-SA is in 330 cases of higher quality than SA. Looking at the individual forecasts , S-AEP is in 324, S-Bays in 354, S-Holt in 322 and S-Quadr in 341 cases better than the corresponding individual forecast.

4. The multivariate case

Here, we consider multivariate forecasts for a vector of variables which is described in the following.

Situation **(S2)**: Let $\mathbf{F}_{i,T+L} := \left(F_{i,T+L}^{(1)},...,F_{i,T+L}^{(k)}\right)'$, $i = 1,...,n$, be unbiased forecasts for $\mathbf{Y}_{T+L} := \left(Y_{T+L}^{(1)},...,Y_{T+L}^{(k)}\right)'$ at time T, $k \in \mathbb{N}$, $k \geq 2$. We have $E(\mathbf{u}_{i,T+L}) = \mathbf{0}$, where $\mathbf{u}_{i,T+L} := \left(u_{i,T+L}^{(1)},...,u_{i,T+L}^{(k)}\right)'$ and $u_{i,T+L}^{(j)} := Y_{T+L}^{(j)} - F_{i,T+L}^{(j)}$, $i = 1,...,n$, $j = 1,...,k$. Further, we assume that $Cov\left(u_{i,T+L}^{(j)}, Y_{T+L}^{(m)}\right) = 0$, $i = 1,...,n$, $j, m = 1,...,k$ and $Cov(\mathbf{u}_{T+L}) =: \Omega$ is p.d, where $\mathbf{u}_{T+L} := \left(\mathbf{u}_{1,T+L}{}',...,\mathbf{u}_{n,T+L}{}'\right)'$. Finally, there exists a vector $\mathbf{u}_{i,T+L}$, without loss of generality $\mathbf{u}_{i,T+L} = \mathbf{u}_{n,T+L}$, so that $Cov[\left((\mathbf{u}_{1,T+L} - \mathbf{u}_{n,T+L})',\ldots,(\mathbf{u}_{n-1,T+L} - \mathbf{u}_{n,T+L})'\right)']$ is p.d.

We consider forecast combinations of the form $\mathbf{F}_{C,T+L} := \sum_{i=1}^{n} \mathbf{C}_i \mathbf{F}_{i,T+L}$, where $\mathbf{C}_i \in \mathbb{R}^{k \times k}$. The MMSE-optimal unbiased forecast combination, where $\sum_{i=1}^{n} \mathbf{C}_i = \mathbf{I}_k$, is given by the weights (see Wenzel, 1998)

$$\mathbf{C}_{opt,unb} := [\mathbf{C}_{1,opt,unb},...,\mathbf{C}_{n,opt,unb}] := [\mathbf{W}'\mathbf{V}^{-1}, \mathbf{I}_k - \mathbf{W}'\mathbf{V}^{-1}\mathbf{I}_k^*] \tag{6}$$

where

$\Omega := (\Omega_{rs})_{r,s=1,...,n} \sim n \cdot k \times n \cdot k$

$\mathbf{V} := (\mathbf{V}_{rs})_{r,s=1,...,n-1} \sim (n-1)k \times (n-1)k$,

$\mathbf{V}_{rs} := \Omega_{rs} + \Omega_{nn} - \Omega_{rn} - \Omega_{ns}$, $r, s = 1,...,n-1$,

$\mathbf{I}_k^* := [\mathbf{I}_k,...,\mathbf{I}_k]' \sim (n-1)k \times k$,

$\mathbf{W} := (\mathbf{w}_1,...,\mathbf{w}_k) \sim (n-1)k \times k$,

$\mathbf{w}_j := \left(\mathbf{w}_{j1}{}',...,\mathbf{w}_{j,n-1}{}'\right)' \sim (n-1)k \times 1,\ j = 1,...,k$,

$\mathbf{w}_{ji} := (\Omega_{nn} - \Omega_{in})\mathbf{e}_j \sim k \times 1,\ i = 1,...,n-1,\ j = 1,...,k$,

and $\mathbf{e}_j$ denotes the j-th unit vector.

As in the univariate case we now want to calculate the MMSE-optimal weights resulting in general in a biased forecast combination. Therefore, we define $\tilde{\mathbf{F}}_{T+L} := \left(\mathbf{F}_{1,T+L}{}',...,\mathbf{F}_{n,T+L}{}'\right)'$,

$\tilde{\mathbf{Y}}_{T+L} := \left(\mathbf{Y}_{T+L}{}',...,\mathbf{Y}_{T+L}{}'\right)' \sim (n \cdot k) \times 1$ and further $\tilde{\mathbf{I}}_k^* := \frac{1}{n}[\mathbf{I}_k,...,\mathbf{I}_k]' \sim n \cdot k \times k$ so that $\mathbf{Y}_{T+L} = \tilde{\mathbf{I}}_k^{*\prime}\tilde{\mathbf{Y}}_{T+L}$.

<u>*Theorem 3*</u>: Considering the situation **(S2)**, the MMSE-optimal forecast combination of the n individual forecasts $\mathbf{F}_{1,T+L},...,\mathbf{F}_{n,T+L}$ is given by $\mathbf{F}_{\mathbf{C}_{opt,b},T+L} := \mathbf{C}_{opt,b}\tilde{\mathbf{F}}_{T+L}$, where

$$\mathbf{C}_{opt,b} = \tilde{\mathbf{I}}_k^{*\prime} E\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\left(\Omega + E\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)^{-1}.$$

Proof: We get

$$\begin{aligned}\mathbf{u}_{\mathbf{C},T+L} &:= \mathbf{Y}_{T+L} - \mathbf{F}_{\mathbf{C},T+L} = \tilde{\mathbf{I}}_k^{*\prime}\tilde{\mathbf{Y}}_{T+L} - \mathbf{C}\tilde{\mathbf{F}}_{T+L} \\ &= \mathbf{C}\mathbf{u}_{T+L} - \left(\mathbf{C} - \tilde{\mathbf{I}}_k^{*\prime}\right)\tilde{\mathbf{Y}}_{T+L},\end{aligned}$$

where $\mathbf{C} := [\mathbf{C}_1,...,\mathbf{C}_n]$ and

$$E\left(\mathbf{u}_{\mathbf{C},T+L}\right) = -\left(\mathbf{C} - \tilde{\mathbf{I}}_k^{*\prime}\right)E\left(\tilde{\mathbf{Y}}_{T+L}\right),$$

$$Cov\left(\mathbf{u}_{\mathbf{C},T+L}\right) = \mathbf{C}\Omega\mathbf{C}' + \left(\mathbf{C} - \tilde{\mathbf{I}}_k^{*\prime}\right)Cov\left(\tilde{\mathbf{Y}}_{T+L}\right)\left(\mathbf{C} - \tilde{\mathbf{I}}_k^{*\prime}\right)'$$

which gives us

$$\begin{aligned}&MMSE\left(\mathbf{Y}_{T+L},\mathbf{F}_{\mathbf{C},T+L}\right) = \mathbf{C}\Omega\mathbf{C}' + \left(\mathbf{C} - \tilde{\mathbf{I}}_k^{*\prime}\right)E\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\left(\mathbf{C} - \tilde{\mathbf{I}}_k^{*\prime}\right)' \\ &= \mathbf{C}\left(\Omega + E\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)\mathbf{C}' - \tilde{\mathbf{I}}_k^{*\prime}E\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}' - \mathbf{C}E\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\tilde{\mathbf{I}}_k^* + \tilde{\mathbf{I}}_k^{*\prime}E\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\tilde{\mathbf{I}}_k^*.\end{aligned}$$

At first we minimize the SMSE and calculate

$$\frac{\partial\, tr\left(MMSE\left(\mathbf{Y}_{T+L},\mathbf{F}_{\mathbf{C},T+L}\right)\right)}{\partial\mathbf{C}} = 2\mathbf{C}\left(\Omega + E\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right) - 2\tilde{\mathbf{I}}_k^{*\prime}E\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right) \overset{!}{=} \mathbf{0}_{k\times(n\cdot k)}$$

$$\Leftrightarrow \mathbf{C}_{opt,b} = \tilde{\mathbf{I}}_k^{*\prime}E\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\left(\Omega + E\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)^{-1}$$

and

$$\frac{\partial^2\, tr\left(MMSE\left(\mathbf{Y}_{T+L},\mathbf{F}_{\mathbf{C},T+L}\right)\right)}{\partial^2\mathbf{C}} = 2\left(\Omega + E\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right) \text{ is p.d.}$$

A forecast combination with an arbritary weight matrix $\mathbf{C}_{arb,b}$, which can be expressed by $\mathbf{C}_{arb,b} := \mathbf{K}\left(\Omega + E\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)^{-1}$, $\mathbf{K} \in \mathbb{R}^{k\times(n\cdot k)}$, cannot outperform $\mathbf{F}_{\mathbf{C}_{opt,b},T+L}$, since

$$\mathrm{MMSE}\left(\mathbf{Y}_{T+L},\mathbf{F}_{\mathbf{C}_{arb,b},T+L}\right)-\mathrm{MMSE}\left(\mathbf{Y}_{T+L},\mathbf{F}_{\mathbf{C}_{opt,b},T+L}\right)$$
$$=\left(\mathbf{K}'-\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}'\right)\tilde{\mathbf{I}}_k^*\right)'\left(\Omega+\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}'\right)\right)^{-1}\left(\mathbf{K}'-\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}'\right)\tilde{\mathbf{I}}_k^*\right),$$
which is obviously n.n.d. and thus $\mathrm{MMSE}\left(\mathbf{Y}_{T+L},\mathbf{F}_{\mathbf{C}_{arb,b},T+L}\right)\geq_L \mathrm{MMSE}\left(\mathbf{Y}_{T+L},\mathbf{F}_{\mathbf{C}_{opt,b},T+L}\right)$.

So far we analysed only the MMSE-optimal unbiased (in the class where $\sum_{i=1}^{n}\mathbf{C}_i=\mathbf{I}_k$) and the MMSE-optimal biased combination. In the following we present the shrinkage approach for the multivariate case.

<u>*Theorem 4*:</u> Consider in **(S2)** an unbiased forecast combination $\mathbf{F}_{\mathbf{C},T+L}:=\sum_{i=1}^{n}\mathbf{C}_i\mathbf{F}_{i,T+L}$, where $\sum_{i=1}^{n}\mathbf{C}_i=\mathbf{I}_k$. A forecast combination $\mathbf{F}_{\lambda\mathbf{C},T+L}:=\lambda\mathbf{F}_{\mathbf{C},T+L}$, $\lambda\in\mathbb{R}$, improves $\mathbf{F}_{\mathbf{C},T+L}$ in the sense of the SMSE for $\lambda\in\left(\dfrac{\mathrm{tr}\left\{\mathbf{C}\left(\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}'\right)-\Omega\right)\mathbf{C}'\right\}}{\mathrm{tr}\left\{\mathbf{C}\left(\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}'\right)+\Omega\right)\mathbf{C}'\right\}},1\right)$. The forecast combination $\mathbf{F}_{\lambda\mathbf{C},T+L}$ with minimal SMSE is given by $\lambda_{\mathbf{C},opt}=\dfrac{\mathrm{tr}\left(\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}'\right)\mathbf{C}'\right)}{\mathrm{tr}\left\{\mathbf{C}\left(\Omega+\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}'\right)\right)\mathbf{C}'\right\}}$.

Proof: We have

$$\mathbf{u}_{\lambda\mathbf{C},T+L}:=\mathbf{Y}_{T+L}-\mathbf{F}_{\lambda\mathbf{C},T+L}=\mathbf{C}\tilde{\mathbf{Y}}_{T+L}-\lambda\mathbf{C}\tilde{\mathbf{F}}_{T+L}$$
$$=\lambda\mathbf{C}\tilde{\mathbf{Y}}_{T+L}-\lambda\mathbf{C}\tilde{\mathbf{F}}_{T+L}-(\lambda-1)\mathbf{C}\tilde{\mathbf{Y}}_{T+L}=\lambda\mathbf{C}\mathbf{u}_{T+L}-(\lambda-1)\mathbf{C}\tilde{\mathbf{Y}}_{T+L},$$

where $\mathbf{C}:=[\mathbf{C}_1,...,\mathbf{C}_n]$.

Hence,

$$\mathrm{E}\left(\mathbf{u}_{\lambda\mathbf{C},T+L}\right)=(1-\lambda)\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\right)$$

and

$$\mathrm{Cov}\left(\mathbf{u}_{\lambda\mathbf{C},T+L}\right)=\lambda^2\mathbf{C}\Omega\mathbf{C}'+(\lambda-1)^2\mathbf{C}\mathrm{Cov}\left(\tilde{\mathbf{Y}}_{T+L}\right)\mathbf{C}'$$

resulting in

$$\mathrm{MMSE}\left(\mathbf{Y}_{T+L},\mathbf{F}_{\lambda\mathbf{C},T+L}\right)=\lambda^2\mathbf{C}\Omega\mathbf{C}'+(\lambda-1)^2\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}'\right)\mathbf{C}' \quad .$$

As in the univariate case we first have a look at the improvement region in the sense of the SMSE. We compare $\mathrm{SMSE}\left(\mathbf{Y}_{T+L},\mathbf{F}_{\lambda\mathbf{C},T+L}\right)$ with $\mathrm{SMSE}\left(\mathbf{Y}_{T+L},\mathbf{F}_{\mathbf{C},T+L}\right)$: For which λ's does

$$\lambda^2\mathrm{tr}\left(\mathbf{C}\left(\Omega+\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)\mathbf{C}'\right)-2\lambda\,\mathrm{tr}\left(\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\right)+\mathrm{tr}\left(\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\right)<\mathrm{tr}\left(\mathbf{C}\Omega\mathbf{C}'\right)$$

hold?

The left side is a quadratic function in λ. Similar to the univariate case we can conclude

$$\mathrm{IR}_{\mathbf{F}_{\mathbf{C},T+L}}:=\left(\frac{\mathrm{tr}\left\{\mathbf{C}\left(\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)-\Omega\right)\mathbf{C}'\right\}}{\mathrm{tr}\left\{\mathbf{C}\left(\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)+\Omega\right)\mathbf{C}'\right\}},1\right)\quad. \tag{7}$$

Since $\mathrm{rg}(\mathbf{C})=k$ (see Appendix), we derive

$$\frac{\partial\,\mathrm{tr}\left(\mathrm{MMSE}\left(\mathbf{Y}_{T+L},\mathbf{F}_{\lambda\mathbf{C},T+L}\right)\right)}{\partial\lambda}=2\lambda\,\mathrm{tr}\left(\mathbf{C}\Omega\mathbf{C}'\right)+2(\lambda-1)\mathrm{tr}\left(\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\right)\overset{!}{=}0$$

$$\Leftrightarrow\lambda_{\mathbf{C},opt}=\frac{\mathrm{tr}\left(\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\right)}{\mathrm{tr}\left\{\mathbf{C}\left(\Omega+\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)\mathbf{C}'\right\}}\quad.$$

Since $\dfrac{\partial^2\mathrm{tr}\left(\mathrm{MMSE}\left(\mathbf{Y}_{T+L},\mathbf{F}_{\lambda\mathbf{C},T+L}\right)\right)}{\partial^2\lambda}=2\mathrm{tr}\left(\mathbf{C}\left(\Omega+\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)\mathbf{C}'\right)>0$, $\lambda_{\mathbf{C},opt}$ is the minimizing scalar.

As in the univariate case, $\lambda_{\mathbf{C},opt}$ is the midpoint of the improvement region. The SMSE of a forecast combination $\lambda_{\mathbf{C},opt}\mathbf{F}_{\mathbf{C},T+L}$ is:

$$\mathrm{SMSE}\left(\mathbf{Y}_{T+L},\lambda_{\mathbf{C},opt}\mathbf{F}_{\mathbf{C},T+L}\right)=\frac{\mathrm{tr}\left(\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\right)\mathrm{tr}\left(\mathbf{C}\Omega\mathbf{C}'\right)}{\mathrm{tr}\left(\mathbf{C}\left(\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)+\Omega\right)\mathbf{C}'\right)}\quad.$$

In general we use an n.n.d estimator for $\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)$. Looking at the given interval in (7), we can see again that an improvement is only possible for $\lambda<1$. Furthermore, we stay in the improvement region if

$$\mathrm{tr}\left(\mathbf{C}\hat{\mathrm{E}}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\right) > \frac{\mathrm{tr}\left(\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\right)}{2} - \frac{\mathrm{tr}(\mathbf{C}\Omega\mathbf{C}')}{2} \qquad (\Omega \text{ known}), \text{ or}$$

$$\mathrm{tr}\left(\mathbf{C}\hat{\mathrm{E}}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\right) > \frac{\mathrm{tr}\left(\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\right)\mathrm{tr}(\mathbf{C}\hat{\Omega}\mathbf{C}')}{2\mathrm{tr}(\mathbf{C}\Omega\mathbf{C}')} - \frac{\mathrm{tr}(\mathbf{C}\hat{\Omega}\mathbf{C}')}{2} \qquad (\Omega \text{ unknown}),$$

where $\hat{\Omega}$ denotes a p.d. estimator of Ω and $\hat{\mathrm{E}}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)$ is an n.n.d estimator, both realised. We can see, that any unbiased forecast combination can be outperformed in the sense of the SMSE by shrinking, especially the MMSE-optimal given in (6). Comparing the MMSEs of a given multivariate forecast and its shrunken version can result in a situation, where none of the MMSEs dominates the other. The difference of the two MMSEs can be indefinit.

Instead of a shrinkage scalar λ we now use a matrix $\Gamma \in \mathbb{R}^{k\times k}$. We consider $\mathbf{F}_{\Gamma\mathbf{C},T+L} := \Gamma\mathbf{F}_{\mathbf{C},T+L}$.

<u>*Theorem 5*:</u> Consider in situation **(S2)** an unbiased forecast combination $\mathbf{F}_{\mathbf{C},T+L}$, where $\sum_{i=1}^{n}\mathbf{C}_i = \mathbf{I}_k$. The MMSE-optimal shrunken combination is given by $\mathbf{F}_{\Gamma_{\mathbf{C},opt}\mathbf{C},T+L} := \Gamma_{\mathbf{C},opt}\mathbf{F}_{\mathbf{C},T+L}$,

where
$$\Gamma_{\mathbf{C},opt} := \mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\left(\mathbf{C}\left(\Omega + \mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)\mathbf{C}'\right)^{-1}. \qquad (8)$$

Proof: The combined forecast error is

$$\mathbf{u}_{\Gamma\mathbf{C},T+L} := \mathbf{Y}_{T+L} - \mathbf{F}_{\Gamma\mathbf{C},T+L} = \Gamma\mathbf{C}\tilde{\mathbf{Y}}_{T+L} - \Gamma\mathbf{C}\tilde{\mathbf{F}}_{T+L} - (\Gamma\mathbf{C} - \mathbf{C})\tilde{\mathbf{Y}}_{T+L}$$
$$= \Gamma\mathbf{C}\mathbf{u}_{T+L} - (\Gamma\mathbf{C} - \mathbf{C})\tilde{\mathbf{Y}}_{T+L},$$

where $\mathbf{C} := [\mathbf{C}_1, ..., \mathbf{C}_n]$,

and

$$\mathrm{E}(\mathbf{u}_{\Gamma\mathbf{C},T+L}) = -(\Gamma\mathbf{C} - \mathbf{C})\mathrm{E}(\tilde{\mathbf{Y}}_{T+L}),$$

$$\mathrm{Cov}(\mathbf{u}_{\Gamma\mathbf{C},T+L}) = \Gamma\mathbf{C}\Omega\mathbf{C}'\Gamma' + (\Gamma\mathbf{C} - \mathbf{C})\mathrm{Cov}(\tilde{\mathbf{Y}}_{T+L})(\Gamma\mathbf{C} - \mathbf{C})'.$$

Calculating the corresponding MMSE results in

$$\mathrm{MMSE}(\mathbf{Y}_{T+L}, \mathbf{F}_{\Gamma\mathbf{C},T+L}) = \Gamma\mathbf{C}\Omega\mathbf{C}'\Gamma' + \Gamma\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\Gamma' - \mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\Gamma'$$
$$- \Gamma\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}' + \mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'.$$

Here we want to minimize $\mathrm{MMSE}\left(\mathbf{Y}_{T+L}, \mathbf{F}_{\Gamma\mathbf{C},T+L}\right)$ with respect to Γ.

$$\frac{\partial\, \mathrm{tr}\left(\mathrm{MMSE}\left(\mathbf{Y}_{T+L}, \mathbf{F}_{\Gamma\mathbf{C},T+L}\right)\right)}{\partial \Gamma} = 2\Gamma\mathbf{C}\left(\Omega + \mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)\mathbf{C}' - 2\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}' \overset{!}{=} \mathbf{0}_{k\times k}$$

$$\Leftrightarrow \Gamma\left(\mathbf{C}\left(\Omega + \mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)\mathbf{C}'\right) = \mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'$$

Since $\mathrm{rg}(\mathbf{C}) = k$ (see Appendix), we get:

$$\Gamma_{\mathbf{C},\mathrm{opt}} := \mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\left(\mathbf{C}\left(\Omega + \mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)\mathbf{C}'\right)^{-1} \quad ,$$

and

$$\frac{\partial^2\, \mathrm{tr}\left(\mathrm{MMSE}\left(\mathbf{Y}_{T+L}, \mathbf{F}_{\Gamma\mathbf{C},T+L}\right)\right)}{\partial^2 \Gamma} = 2\mathbf{C}\left(\Omega + \mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)\mathbf{C}'$$

is p.d. For an arbitrary shrinkage matrix $\Gamma_{\mathbf{C},\mathrm{arb}} := \mathbf{K}\left(\mathbf{C}\left(\Omega + \mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)\mathbf{C}'\right)^{-1}$, $\mathbf{K} \in \mathbb{R}^{k\times k}$,

we get

$$\mathrm{MMSE}\left(\mathbf{Y}_{T+L}, \mathbf{F}_{\Gamma_{\mathbf{C},\mathrm{arb}}\mathbf{C},T+L}\right) - \mathrm{MMSE}\left(\mathbf{Y}_{T+L}, \mathbf{F}_{\Gamma_{\mathbf{C},\mathrm{opt}}\mathbf{C},T+L}\right)$$

$$= \left(\mathbf{K}' - \mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\right)'\left(\mathbf{C}\left(\Omega + \mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\right)\mathbf{C}'\right)^{-1}\left(\mathbf{K}' - \mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\tilde{\mathbf{Y}}_{T+L}{}'\right)\mathbf{C}'\right),$$

which is n.n.d and thus $\mathrm{MMSE}\left(\mathbf{Y}_{T+L}, \mathbf{F}_{\Gamma_{\mathbf{C},\mathrm{arb}}\mathbf{C},T+L}\right) \geq_L \mathrm{MMSE}\left(\mathbf{Y}_{T+L}, \mathbf{F}_{\Gamma_{\mathbf{C},\mathrm{opt}}\mathbf{C},T+L}\right)$

We cannot see directly from the form of the weights of the MMSE-optimal unbiased and of the MMSE-optimal (biased) combination how they are related. But looking also at the results of the study in section 5 shows us that $\mathbf{C}_{\mathrm{opt,b}} = \Gamma_{\mathbf{C}_{\mathrm{opt,unb}},\mathrm{opt}}\mathbf{C}_{\mathrm{opt,unb}}$. In the simulation study we only analyse forecast combinations where the weight matrices sum up to $\mathbf{I}_k$. This is not a necessary condition for an unbiased forecast combination. We can also demand $\mathbf{C}\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\right) = \tilde{\mathbf{I}}_k^{*\prime}\,\mathrm{E}\left(\tilde{\mathbf{Y}}_{T+L}\right)$. In that case the weights depend on $\mathrm{E}\left(\mathbf{Y}_{T+L}\right)$. Estimating this results in general in weights which gives us a biased forecast combination. Then the unbiasedness assumption in Thereoms 5 and 6 is not valid. Thus we do not consider these techniques.

5. Simulation study for the multivariate case

We analyse the combination of three unbiased one-step individual forecasts for a two-dimensional variable. We use $E(\mathbf{Y}_t) = (5,5)'$, $t = 1,...,30$, for the generation of the time series of the variable to be forecasted. For $\mathrm{Cov}(\mathbf{Y}_t) =: \Lambda_i, i = 1,2$, $t = 1,..,30$ we consider:

$$\Lambda_1 := \begin{pmatrix} 19 & 9 \\ 9 & 30 \end{pmatrix}, \ \Lambda_2 := \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}.$$

For the generation of the series of the forecast errors we use 20 different 6×6 covariance matrices which are given in the Appendix. The time series of the **Y**-variable and of the forecast errors are of length 30. We have two different covariance matrices of **Y** and 20 different error covariance matrices. This results in 40 different cases. For each case we generate 100 times series for **Y** and for the individual forecast errors.

The first combination weights of the different methods are calculated on the basis of the first ten data points. Thus, 20 data points are left for our analysis. In each step the different unknown parameters are re-estimated on the basis of the most recent 10 data points. We analyse the following 15 techniques:

T1: MMSE-optimal unbiased combination given by equation (6)

T2: MMSE-optimal (biased) combination given in Theorem 3

T3: shrinking T1 with the corresponding optimal shrinkage scalar λ

T4: simple average (SA) of the individual forecast

T5: shrinking SA with the corresponding optimal shrinkage scalar λ

T6: shrinking SA with the corresponding optimal shrinkage matrix Γ

T7: individual forecast No. 1

T8: individual forecast No. 2

T9: individual forecast No. 3

T10: shrinking individual forecast No. 1 with the corresponding optimal shrinkage scalar λ

T11: shrinking individual forecast No. 2 with the corresponding optimal shrinkage scalar λ

T12: shrinking individual forecast No. 3 with the corresponding optimal shrinkage scalar λ

T13: shrinking individual forecast No. 1 with the corresponding optimal shrinkage matrix Γ

T14: shrinking individual forecast No. 2 with the corresponding optimal shrinkage matrix Γ

T15: shrinking individual forecast No. 3 with the corresponding optimal shrinkage matrix Γ.

Again, we want to remark that shrinking T1 with the corresponding optimal shrinkage matrix Γ is identical to T2. In the following tables we present the average of the MSEs (first value in the tables) of the 100 time series in each case for both components. We also count in each case for how many time series a certain combination technique performs better than the simple average of the individual forecasts (second value in the tables).

Table 2: Results for component No. 1, $\mathrm{Cov}(\mathbf{Y}_t) = \Lambda_1$

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
Ω_1	7.07 82	6.27 89	6.71 85	10.07 -	8.53 88	7.99 82	14.20 16	27.41 3	43.30 0	11.79 35	18.28 8	23.81 1	11.31 44	14.34 22	17.22 11
Ω_2	2.72 72	2.65 72	2.61 70	3.45 -	3.22 66	3.37 52	12.72 0	18.61 0	6.96 3	10.23 0	13.80 0	6.39 5	6.35 10	13.56 0	6.60 2
Ω_3	1.49 66	1.43 70	1.45 65	1.71 -	1.67 60	1.64 57	5.15 1	7.19 0	6.04 0	4.78 1	6.38 0	5.48 0	4.90 2	6.30 0	5.53 0
Ω_4	3.90 93	3.52 96	3.76 93	7.64 -	6.81 80	5.99 84	13.02 2	13.30 1	15.26 2	10.78 18	10.95 15	12.65 10	10.18 22	10.91 16	11.37 16
Ω_5	2.10 58	1.99 64	2.05 60	2.42 -	2.45 48	2.48 47	11.93 0	6.76 0	10.44 0	10.10 0	6.10 0	9.08 0	8.53 1	6.28 0	6.20 4
Ω_6	2.05 85	2.00 88	1.98 89	3.08 -	3.37 33	3.18 43	6.96 4	14.44 0	8.12 2	6.41 3	11.90 0	7.23 3	6.62 3	11.74 0	5.60 7
Ω_7	1.75 79	1.68 79	1.71 79	2.32 -	2.25 62	2.24 63	3.96 8	6.85 3	15.81 0	3.72 12	6.33 3	12.01 0	3.24 24	5.31 5	11.23 0
Ω_8	2.49 99	2.43 99	2.42 99	5.31 -	4.82 77	4.91 68	18.99 1	37.43 0	17.66 0	14.47 2	21.33 0	13.66 0	14.04 2	11.32 1	11.95 2
Ω_9	3.27 70	3.19 70	3.18 71	3.79 -	3.59 71	3.80 57	5.10 25	13.55 0	3.93 43	4.80 29	10.94 0	3.69 58	4.82 30	10.78 0	3.87 48
Ω_{10}	1.29 66	1.02 86	1.28 69	1.49 -	1.82 26	1.38 58	9.73 0	9.88 0	2.95 2	8.46 0	8.22 0	2.81 3	7.84 0	7.07 0	1.77 35
Ω_{11}	2.12 93	2.10 93	2.08 93	3.70 -	3.41 75	3.51 60	8.88 2	12.83 0	24.85 0	7.75 6	10.84 1	16.26 0	8.08 6	10.79 0	15.85 0
Ω_{12}	0.84 100	0.80 100	0.82 100	2.56 -	2.46 62	2.41 64	4.27 11	25.16 0	3.05 35	3.97 14	17.06 0	2.95 40	3.43 30	13.60 0	2.85 42
Ω_{13}	2.53 90	2.36 92	2.46 91	3.89 -	3.58 65	3.72 62	16.55 0	14.22 0	6.93 6	12.93 0	11.17 1	6.16 16	13.38 0	10.89 0	5.69 22
Ω_{14}	5.08 21	4.45 33	4.92 27	3.90 -	3.89 46	3.70 59	16.98 0	8.78 0	17.20 0	13.66 0	7.77 1	13.27 0	14.30 0	6.08 6	9.41 1
Ω_{15}	1.36 82	1.23 89	1.33 82	1.95 -	1.95 46	1.76 70	16.13 0	4.19 2	12.90 0	12.66 0	3.96 5	10.43 0	10.64 0	3.42 10	7.19 0
Ω_{16}	0.57 100	0.57 100	0.57 100	5.30 -	4.86 75	5.13 57	18.54 0	3.00 94	4.93 59	14.32 0	2.88 95	4.66 71	13.69 0	2.89 96	4.63 65
Ω_{17}	1.49 99	1.44 99	1.46 99	3.88 -	3.65 67	3.87 49	24.24 0	6.98 8	9.59 2	16.91 0	6.26 12	8.27 5	13.29 0	6.40 11	6.75 7
Ω_{18}	0.49 100	0.49 100	0.52 100	2.58 -	2.66 42	2.61 51	5.08 8	28.90 0	18.10 0	4.72 11	18.13 0	13.64 0	4.69 14	17.24 0	13.93 0
Ω_{19}	4.85 61	4.56 68	4.62 68	5.71 -	5.19 73	5.55 63	17.91 0	13.79 0	6.00 45	13.90 0	11.48 1	5.50 51	10.19 7	12.07 0	5.74 49
Ω_{20}	2.77 16	2.40 27	2.68 16	1.97 -	1.95 45	1.88 65	7.21 0	11.67 0	4.77 2	6.53 0	9.71 0	4.41 6	6.40 0	10.04 0	4.43 4

Looking at the results for component No. 1 we see that method No. 2 is best. Shrinking the unbiased forecast combinations leads to an improvement. Using a shrinkage matrix Γ is in most cases better than the usage of a shrinkage scalar λ. Only for the simple average combination the approach with the shrinkage scalar performs for 9 of the given covariance matrices better. The simple average is also outperformed by the other combinations. Combining leads in general to an improvement of the individual forecasts.

Table 3: Results for component No. 2, $\mathrm{Cov}(\mathbf{Y}_t) = \Lambda_1$

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
Ω_1	0.97 100	0.95 100	0.98 99	4.12 -	4.14 50	4.00 56	3.08 76	8.24 1	17.00 0	5.69 25	11.92 1	18.31 0	3.11 77	7.23 4	11.68 0
Ω_2	5.01 10	4.78 14	4.78 13	3.50 -	3.39 64	3.55 44	15.28 0	8.84 0	8.97 4	12.68 0	9.70 1	8.00 10	7.97 1	8.42 1	8.48 7
Ω_3	3.35 45	3.16 51	3.27 47	3.15 -	3.03 65	3.04 55	18.84 0	8.35 1	7.97 2	15.80 0	7.74 1	7.34 3	14.06 0	7.73 1	7.62 5
Ω_4	2.59 58	2.41 66	2.52 61	2.79 -	3.00 39	2.42 74	10.27 1	12.27 0	24.37 0	10.22 1	10.89 0	18.29 0	8.62 2	10.48 0	16.32 0
Ω_5	2.13 100	2.02 100	2.07 100	6.39 -	5.92 73	6.22 53	13.16 1	20.56 0	7.60 33	11.28 4	16.73 0	7.85 31	9.88 9	15.84 0	4.88 73
Ω_6	3.60 100	3.49 99	3.48 100	11.61 -	10.05 88	10.29 72	25.25 0	12.76 43	26.21 0	20.38 5	11.96 46	20.22 0	17.52 13	11.67 48	13.92 31
Ω_7	2.06 96	1.97 97	2.01 97	4.76 -	4.44 70	4.47 63	7.91 4	10.66 0	2.91 100	7.25 7	9.06 2	6.04 35	6.12 25	8.06 6	2.99 100
Ω_8	3.10 71	2.99 77	2.99 77	3.75 -	3.70 54	3.58 56	16.22 0	23.11 0	11.75 0	14.12 0	20.35 0	11.76 0	13.44 0	8.76 1	9.30 3
Ω_9	3.62 71	3.49 75	3.48 76	4.41 -	4.08 72	4.35 49	9.08 0	10.52 3	12.62 0	7.92 0	9.54 3	10.95 2	8.01 2	9.43 4	10.88 4
Ω_{10}	9.06 63	7.02 89	8.55 66	10.51 -	9.19 85	7.98 88	8.54 63	27.45 0	20.65 0	8.47 71	20.57 0	18.21 0	7.35 80	15.50 16	10.87 47
Ω_{11}	3.81 65	3.62 70	3.66 68	4.31 -	4.02 68	4.16 59	19.94 0	15.11 0	9.04 5	15.87 0	12.50 0	11.18 2	15.36 1	12.40 0	8.39 5
Ω_{12}	1.64 85	1.58 88	1.63 87	2.64 -	2.52 67	2.53 62	7.36 2	8.51 0	4.94 6	6.70 0	11.73 0	4.58 11	5.86 4	7.24 0	4.53 11
Ω_{13}	0.76 100	0.75 100	0.76 100	4.67 -	4.44 72	4.61 58	10.16 2	16.26 0	6.98 13	10.47 4	13.49 0	6.81 19	9.56 5	13.33 0	6.41 23
Ω_{14}	0.36 100	0.32 100	0.38 100	10.49 -	8.98 78	8.42 75	44.05 0	8.74 74	4.87 99	28.06 0	8.18 74	7.59 78	26.63 1	6.19 95	3.43 99
Ω_{15}	1.58 99	1.44 99	1.56 99	4.63 -	4.33 79	4.04 74	25.83 0	19.27 0	5.05 45	19.06 0	16.30 0	6.42 26	14.86 0	11.60 0	3.13 80
Ω_{16}	1.80 59	1.78 59	1.78 59	1.87 -	1.98 39	1.91 45	9.32 0	2.83 15	5.61 0	10.55 0	2.81 15	5.29 0	8.78 0	2.81 15	5.29 1
Ω_{17}	2.15 93	2.06 93	2.11 93	3.76 -	3.61 57	3.88 45	2.07 92	6.67 1	12.10 0	8.75 7	6.52 2	10.67 0	1.94 95	6.76 0	8.60 2
Ω_{18}	8.13 61	7.23 70	7.68 65	8.90 -	7.83 83	8.16 68	9.63 39	26.21 0	23.78 0	8.29 53	19.71 2	17.92 3	7.98 61	18.44 0	18.38 2
Ω_{19}	2.49 94	2.34 96	2.42 96	5.31 -	4.96 63	5.16 56	14.00 4	15.64 0	24.11 0	12.85 2	13.01 1	19.79 0	8.94 16	13.44 1	16.55 0
Ω_{20}	2.17 93	1.88 97	2.13 93	4.35 -	4.17 63	4.10 62	15.87 0	9.94 0	10.83 0	13.21 0	9.91 0	9.50 2	12.47 0	9.47 1	9.12 5

Again, method No. 2 is best. Combining the forecasts in most cases leads to an improvement. Using a shrinkage matrix Γ is in general better than using a scalar λ. Only for the simple average combination we have similar results as above.

To summarize the results we present the following table. The first value in the first row gives us the number of the 20 cases, where the average MSE of the first component of the special shrinked combination is smaller than that of the corresponding unbiased forecast combination. The first numbers of the second row are the same for the second component. The first numbers of the third row presents how often the sum of the two averages of the MSEs is smaller than that of the corresponding unbiased combination. We also count how often the special methods are best (*second numbers*).

Table 4: Summary of the results, $\mathrm{Cov}(\mathbf{Y}_t) = \Lambda_1$

	T2	T3	T5	T6	T10	T13	T11	T14	T12	T15
component No. 1	20 *13*	19 *5*	15 *0*	16 *2*	19 *0*	20 *0*	20 *0*	20 *0*	20 *0*	20 *0*
component No. 2	20 *15*	17 3	17 *2*	17 *0*	16 *0*	17 *0*	13 *0*	19 *0*	19 *0*	19 *0*
sum of av. MSEs	20 *16*	20 *3*	20 *0*	19 *1*	19 *0*	20 *0*	20 *0*	20 *0*	20 *0*	20 *0*

Again, shrinking the unbiased forecasts leads to an improvement. For some covariance matrices the unbiased forecast combinations are for a special component better than their shrinkage versions. But looking at the sum of averaged MSEs, in almost all cases the shrinkage techniques are better.

As above, we want to describe now the result for $\mathrm{Cov}(\mathbf{Y}_t) = \Lambda_2$.

Table 5: Results for component No. 1, $\mathrm{Cov}(\mathbf{Y}_t) = \Lambda_2$

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
Ω_1	7.06 80	5.45 91	6.61 80	9.67 -	7.71 92	5.32 99	13.94 13	26.86 2	43.01 0	10.03 46	15.74 9	19.41 2	8.05 65	8.91 54	10.98 38
Ω_2	2.50 80	2.48 82	2.37 88	3.59 -	3.32 71	3.40 63	12.72 1	20.11 0	7.05 6	9.83 3	12.73 1	6.00 10	3.48 48	11.30 0	6.22 9
Ω_3	1.55 60	1.37 75	1.49 67	1.74 -	1.75 52	1.30 89	4.93 2	7.01 0	5.90 0	4.43 3	6.08 0	5.29 0	4.48 5	5.07 0	4.92 0
Ω_4	4.19 88	2.68 99	3.88 92	7.78 -	6.47 87	3.45 98	13.17 0	13.01 1	14.91 4	9.98 22	9.82 23	10.69 19	7.78 51	9.14 33	9.15 31
Ω_5	1.97 72	1.69 82	1.90 73	2.48 -	2.55 42	2.49 48	12.24 0	7.15 0	10.66 0	9.51 0	6.25 2	8.30 1	6.44 0	6.49 2	2.73 44
Ω_6	1.96 91	1.91 92	1.86 93	3.23 -	3.48 34	3.20 44	6.89 0	14.42 0	8.23 0	5.99 5	10.26 0	6.82 2	6.08 6	9.78 0	4.55 22
Ω_7	1.84 74	1.66 80	1.79 75	2.26 -	2.24 49	2.10 59	3.94 5	6.88 1	15.91 0	3.55 10	6.01 2	11.12 0	2.34 44	4.24 9	7.73 1
Ω_8	2.44 97	2.36 97	2.34 97	5.45 -	4.80 77	3.84 92	18.44 1	36.79 0	16.67 0	11.97 8	17.95 0	11.73 1	10.94 7	5.46 52	8.16 16
Ω_9	3.48 61	3.28 65	3.28 68	3.78 -	3.51 65	3.59 59	5.15 21	13.10 0	3.94 43	4.57 29	10.14 0	3.68 56	4.44 30	8.98 1	3.91 44

Table 5 continiued															
Ω_{10}	1.32 63	0.61 98	1.31 63	1.44 -	2.06 11	1.07 89	9.43 0	9.49 0	2.99 0	7.64 0	7.69 0	2.77 1	5.85 0	5.82 0	0.84 91
Ω_{11}	2.07 89	1.99 90	1.96 93	3.76 -	3.38 73	3.22 71	9.40 3	13.73 0	24.69 0	7.36 6	10.11 3	14.64 0	7.69 8	9.59 2	12.10 0
Ω_{12}	0.85 100	0.73 100	0.84 100	2.38 -	2.29 62	1.79 91	3.91 12	25.28 0	3.11 28	3.62 18	15.24 0	2.92 33	2.24 56	8.57 0	2.32 56
Ω_{13}	2.55 82	2.22 91	2.47 84	3.60 -	3.34 64	3.24 74	15.23 0	13.46 0	7.37 2	10.86 1	9.82 1	6.20 7	10.33 1	8.99 1	4.43 32
Ω_{14}	4.86 32	2.87 83	4.61 34	3.92 -	3.99 40	3.25 76	17.63 0	8.77 0	16.61 0	11.70 1	7.27 5	11.60 0	12.33 0	3.59 59	4.32 41
Ω_{15}	1.36 80	1.00 93	1.34 81	1.98 -	1.96 48	1.28 93	15.85 0	4.03 2	12.77 0	11.09 0	3.71 5	9.35 0	8.44 0	2.83 18	3.14 15
Ω_{16}	0.56 100	0.54 100	0.55 100	5.58 -	4.82 82	4.64 74	19.81 0	3.05 93	4.81 69	12.78 0	2.78 96	4.25 80	10.34 4	2.54 95	3.78 94
Ω_{17}	1.56 100	1.43 100	1.51 100	4.00 -	3.64 71	3.75 65	24.36 0	6.81 11	9.57 4	14.84 0	5.91 16	7.71 5	6.42 14	5.73 19	4.69 37
Ω_{18}	0.52 100	0.51 100	0.56 100	2.47 -	2.64 34	2.46 49	5.36 4	29.33 0	18.20 0	4.70 6	16.00 0	12.32 0	4.28 10	13.88 0	12.35 0
Ω_{19}	5.17 60	4.71 69	4.79 73	5.63 -	4.73 79	4.72 77	17.50 0	14.24 0	6.05 47	11.57 2	10.08 6	5.16 62	6.38 34	10.17 4	5.28 58
Ω_{20}	2.88 16	1.71 73	2.74 20	2.06 -	1.98 59	1.58 88	7.52 0	11.99 0	5.08 2	6.26 0	9.17 0	4.44 2	5.66 0	8.90 0	4.00 5

The results for matrix Λ_2 are similar to the results for matrix Λ_1. For component No. 1 method No. 2 is best. Shrinking the unbiased forecast combinations improves the forecast quality in the sense of the MSE. Using a shrinkage matrix Γ is for all techniques better than using a scalar λ. The simple average is the combination method with lowest quality.

Table 6: Results for component No. 2, $\mathrm{Cov}(\mathbf{Y}_t) = \Lambda_2$

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
Ω_1	0.94 100	0.92 100	0.93 100	3.87 -	3.44 67	2.98 85	2.99 73	8.06 0	16.02 0	4.28 42	8.13 2	11.48 0	2.84 77	5.33 20	7.61 1
Ω_2	5.27 7	4.53 22	4.90 16	3.57 -	3.20 75	3.21 73	15.23 0	9.48 1	9.43 3	10.04 0	8.05 1	7.43 4	3.79 41	7.68 2	6.70 6
Ω_3	3.43 47	2.51 80	3.25 52	3.38 -	2.98 87	2.11 99	19.84 0	8.81 0	8.15 1	15.07 0	6.90 3	6.57 7	6.83 2	5.20 18	5.13 16
Ω_4	2.57 63	1.81 92	2.41 68	2.94 -	2.79 58	1.59 99	10.30 0	11.95 0	24.67 0	7.38 0	8.81 1	14.85 0	6.01 4	7.92 0	8.72 1
Ω_5	2.21 99	1.80 99	2.12 100	6.35 -	5.27 94	4.22 96	12.65 2	20.89 0	7.91 28	8.91 17	14.89 0	6.53 42	5.78 65	8.19 24	1.98 100
Ω_6	3.65 99	3.28 99	3.47 99	11.17 -	8.29 96	6.23 96	24.45 1	12.76 33	25.30 0	17.56 7	8.88 68	17.31 2	8.76 74	8.43 80	6.17 94
Ω_7	1.98 97	1.65 99	1.89 97	4.71 -	4.13 86	3.12 94	7.87 6	10.87 0	2.83 100	6.68 10	8.34 4	4.77 56	3.29 81	4.51 56	2.78 100
Ω_8	3.32 58	3.05 69	3.19 61	3.69 -	3.31 69	2.78 85	16.89 0	22.7ß 0	12.35 0	11.42 1	12.87 0	9.21 0	9.84 1	4.33 35	6.56 8
Ω_9	3.52 74	3.14 81	3.31 78	4.37 -	3.83 85	3.57 82	9.04 0	11.31 0	13.15 0	7.41 1	8.37 8	10.75 1	5.28 31	7.26 15	6.57 9
Ω_{10}	8.65 71	3.66 100	7.81 78	10.89 -	8.28 99	3.77 100	8.87 69	28.24 0	20.90 0	6.99 83	18.22 2	17.48 1	4.92 96	6.89 86	3.65 99
Ω_{11}	3.67 80	3.04 85	3.45 83	4.40 -	3.77 86	3.30 85	20.77 0	14.88 0	9.22 5	13.86 1	9.73 1	8.14 2	9.50 6	8.52 4	7.30 12
Ω_{12}	1.66 84	1.33 95	1.61 87	2.39 -	2.19 76	1.70 92	6.95 0	7.98 0	4.86 3	5.78 1	7.90 0	4.30 7	2.83 39	5.32 0	2.90 33

Table 6 continiued															
Ω_{13}	0.73 100	0.72 100	0.72 100	4.60 -	4.05 83	3.71 83	9.86 6	15.87 0	7.03 11	7.73 7	10.91 0	5.86 26	7.74 10	8.69 3	4.11 59
Ω_{14}	0.35 100	0.21 100	0.36 100	10.97 -	8.16 100	5.09 99	46.50 0	9.08 76	4.89 96	22.77 0	6.97 96	5.67 94	14.31 25	3.44 100	2.18 100
Ω_{15}	1.54 98	1.12 99	1.49 98	4.61 -	4.06 86	2.36 100	25.50 0	18.67 0	4.84 40	15.10 0	15.13 0	5.03 40	8.15 12	5.38 36	1.43 99
Ω_{16}	1.82 65	1.72 66	1.77 65	2.07 -	2.01 50	2.02 53	9.06 0	2.98 18	6.34 0	7.77 0	2.70 35	5.35 0	6.96 0	2.40 40	4.02 5
Ω_{17}	2.27 90	1.97 98	2.19 93	3.90 -	3.45 77	3.47 71	2.00 97	7.05 0	12.31 0	5.90 15	5.83 6	8.85 0	1.65 100	5.48 9	4.89 28
Ω_{18}	8.14 52	5.18 86	7.49 59	8.19 -	6.49 96	5.20 97	10.19 32	24.81 0	21.76 0	8.03 52	13.11 10	12.59 11	5.31 80	11.27 12	11.26 20
Ω_{19}	2.67 97	2.43 97	2.49 97	5.26 -	4.40 86	4.23 85	13.62 3	14.89 0	24.70 0	9.40 12	10.01 4	18.35 0	5.28 45	9.23 6	7.57 20
Ω_{20}	2.20 94	1.34 98	2.10 94	4.16 -	3.69 80	2.46 95	14.86 0	9.88 0	10.77 2	11.06 0	7.47 7	8.68 3	6.49 8	7.01 12	5.59 27

Method No. 2 again is best for the second component. Shrinking leads to a smaller MSE. The forecast combinations improve the individual forecasts. For all combination methods the approach with a shrinkage matrix Γ performs better than the approach with a constant λ. Furthermore, only the simple average is in some cases of lower quality than some of the individual forecasts. We now want to summarize the results in the following table.

Table 7: Summary of the results, $\mathrm{Cov}(\mathbf{Y}_t) = \Lambda_2$

	T2	T3	T5	T6	T10	T13	T11	T14	T12	T15
component No. 1	20 *13*	19 *5*	14 *0*	19 *2*	20 *0*	20 *0*	20 *0*	20 *0*	20 *0*	20 *0*
component No. 2	20 *14*	19 *0*	20 *1*	20 *3*	18 *0*	19 *1*	17 *0*	20 *0*	20 *0*	20 *1*
sum of av. MSEs	20 *19*	20 *0*	20 *0*	20 *1*	20 *0*	20 *0*	20 *0*	20 *0*	20 *0*	20 *0*

We can conclude as for matrix Λ_1. Shrinking leads to an improvement. Method No. 2 is obviously the best.

6. Concluding remarks

Giving up the requirement of unbiased forecast combination improves the quality of the combined forecast in the sense of the MSE (uinivariate), the SMSE or the MMSE (multivariate). Especially the shrinkage approach gives information, how unbiased forecast combinations can be improved. Although the optimal shrinkage scalar (matrix) depends on unknown variables, the second moment of the variable to be forecasted (univariate case) and the error covariance matrix, it is possible to calculate shrinkage scalars lying in the improvement region. An example shows us that the shrinkage versions of different unbiased forecast combinations are of high quality. A more detailed analysis of the performance of the estimators of the unknown parameters is necessary. In this case it could be possible to decide when shrinking is useful and which estimators we should rely on.

In practice subjective weighting schemes are often used for the combination of forecasts. Analysts often decide to weight the forecasts in a special relation, depending on some a-priori knowledge. Then they conclude in common that the weights should sum up to one. Shrinking these forecasts combination saves the relation between the weights. The restriction that the weights sum up to one is no more valid. Then, choosing adequate estimators for the unknown parameters can lead to an improvement.

7. Appendix

A1) We show that a $k \times n \cdot k$ matrix $\mathbf{C} := [\mathbf{C}_1, ..., \mathbf{C}_n]$, where $\sum_{i=1}^{n} \mathbf{C}_i = \mathbf{I}_k$, has full row rank:

$$k = \mathrm{rg}(\mathbf{I}_k) = \mathrm{rg}\left(\sum_{i=1}^{n} \mathbf{C}_i\right) = \mathrm{rg}(\mathbf{C}[\mathbf{I}_k, ..., \mathbf{I}_k]') \le \mathrm{rg}(\mathbf{C}).$$

A2) The 20 error covariance matrices in the simulation study:

$$\Omega_1 := \begin{pmatrix} 14 & 4 & 2 & -3 & 4 & 4 \\ 4 & 3 & 5 & -3 & -1 & -1 \\ 2 & 5 & 27 & -2 & -5 & 1 \\ -3 & -3 & -2 & 8 & -3 & 8 \\ 4 & -1 & -5 & -3 & 42 & -8 \\ 4 & -1 & 1 & 8 & -8 & 16 \end{pmatrix}, \quad \Omega_2 := \begin{pmatrix} 13 & -11 & -6 & 2 & 0 & 7 \\ -11 & 16 & 9 & 2 & 3 & -3 \\ -6 & 9 & 19 & 6 & 2 & -6 \\ 2 & 2 & 6 & 9 & -1 & 0 \\ 0 & 3 & 2 & -1 & 7 & 4 \\ 7 & -3 & -6 & 0 & 4 & 9 \end{pmatrix},$$

$$\Omega_3 := \begin{pmatrix} 5 & 1 & -3 & -3 & -1 & 2 \\ 1 & 19 & -7 & 6 & -2 & -5 \\ -3 & -7 & 7 & 0 & 3 & 1 \\ -3 & 6 & 0 & 9 & -3 & -4 \\ -1 & -2 & 3 & -3 & 6 & 1 \\ 2 & -5 & 1 & -4 & 1 & 8 \end{pmatrix}, \quad \Omega_4 := \begin{pmatrix} 13 & 0 & 10 & 4 & 3 & -11 \\ 0 & 10 & -1 & -5 & 1 & -7 \\ 10 & -1 & 13 & 4 & 1 & -13 \\ 4 & -5 & 4 & 12 & -8 & 2 \\ 3 & 1 & 1 & -8 & 15 & -1 \\ -11 & -7 & -13 & 2 & -1 & 24 \end{pmatrix},$$

$$\Omega_5 := \begin{pmatrix} 12 & -3 & 5 & -2 & -5 & 5 \\ -3 & 13 & 2 & 11 & 3 & -1 \\ 5 & 2 & 7 & 3 & -4 & 5 \\ -2 & 11 & 3 & 21 & 7 & -2 \\ -5 & 3 & -4 & 7 & 11 & -8 \\ 5 & -1 & 5 & -2 & -8 & 8 \end{pmatrix}, \ \Omega_6 := \begin{pmatrix} 7 & 2 & -1 & -3 & 1 & -6 \\ 2 & 25 & 15 & -2 & 0 & 15 \\ -1 & 15 & 15 & 6 & -1 & 14 \\ -3 & -2 & 6 & 13 & -2 & 8 \\ 1 & 0 & -1 & -2 & 8 & -7 \\ -6 & 15 & 14 & 8 & -7 & 27 \end{pmatrix},$$

$$\Omega_7 := \begin{pmatrix} 4 & -3 & -3 & 1 & 5 & 0 \\ -3 & 8 & 2 & 3 & 0 & 3 \\ -3 & 2 & 7 & -3 & -5 & -1 \\ 1 & 3 & -3 & 11 & 2 & 5 \\ 5 & 0 & -5 & 2 & 16 & 2 \\ 0 & 3 & -1 & 5 & 2 & 3 \end{pmatrix}, \ \Omega_8 := \begin{pmatrix} 19 & 5 & -14 & 11 & -3 & 1 \\ 5 & 17 & 13 & -7 & 6 & -8 \\ -14 & 13 & 38 & -24 & 4 & -11 \\ 11 & -7 & -24 & 24 & -3 & 6 \\ -3 & 6 & 4 & -3 & 17 & -2 \\ 1 & -8 & -11 & 6 & -2 & 12 \end{pmatrix},$$

$$\Omega_9 := \begin{pmatrix} 5 & 1 & 0 & 1 & 0 & 0 \\ 1 & 9 & 2 & 1 & 2 & 8 \\ 0 & 2 & 14 & 3 & 6 & 1 \\ 1 & 1 & 3 & 11 & 0 & -6 \\ 0 & 2 & 6 & 0 & 4 & 3 \\ 0 & 8 & 1 & -6 & 3 & 13 \end{pmatrix}, \ \Omega_{10} := \begin{pmatrix} 10 & -1 & -6 & 8 & -1 & 4 \\ -1 & 9 & -3 & -1 & -1 & 0 \\ -6 & -3 & 10 & -5 & 2 & -6 \\ 8 & -1 & -5 & 28 & -6 & 21 \\ -1 & -1 & 2 & -6 & 3 & -7 \\ 4 & 0 & -6 & 21 & -7 & 21 \end{pmatrix},$$

$$\Omega_{11} := \begin{pmatrix} 9 & 4 & -4 & -1 & 0 & 2 \\ 4 & 19 & 8 & 1 & -1 & -7 \\ -4 & 8 & 13 & 4 & -2 & -7 \\ -1 & 1 & 4 & 15 & -2 & 4 \\ 0 & -11 & -2 & -2 & 24 & 6 \\ 2 & -7 & -7 & 4 & 6 & 9 \end{pmatrix}, \ \Omega_{12} := \begin{pmatrix} 4 & -3 & 1 & 1 & -1 & 1 \\ -3 & 7 & -10 & 1 & 2 & -3 \\ 1 & -10 & 25 & -2 & -5 & 5 \\ 1 & 1 & -2 & 8 & -1 & 3 \\ -1 & 2 & -5 & -1 & 3 & -1 \\ 1 & -3 & 5 & 3 & -1 & 5 \end{pmatrix},$$

$$\Omega_{13} := \begin{pmatrix} 16 & 7 & -3 & 8 & -1 & 1 \\ 7 & 10 & -8 & -1 & -5 & -3 \\ -3 & -8 & 14 & 3 & 3 & 7 \\ 8 & -1 & 3 & 16 & -1 & 8 \\ -1 & -5 & 3 & -1 & 7 & -1 \\ 1 & -3 & 7 & 8 & -1 & 7 \end{pmatrix}, \ \Omega_{14} := \begin{pmatrix} 18 & 19 & 1 & 3 & -11 & 6 \\ 19 & 46 & -1 & 14 & -16 & 6 \\ 1 & -1 & 9 & -5 & 6 & 0 \\ 3 & 14 & -5 & 9 & -4 & -1 \\ -11 & -16 & 6 & -4 & 17 & -6 \\ 6 & 6 & 0 & -1 & -6 & 5 \end{pmatrix},$$

$$\Omega_{15} := \begin{pmatrix} 16 & -5 & 2 & -2 & -7 & 4 \\ -5 & 26 & 5 & -7 & -3 & 1 \\ 2 & 5 & 4 & -3 & -3 & 2 \\ -2 & -7 & -3 & 18 & -7 & 2 \\ -7 & -3 & -3 & -7 & 13 & -7 \\ 4 & 1 & 2 & 2 & -7 & 5 \end{pmatrix}, \ \Omega_{16} := \begin{pmatrix} 19 & 4 & 4 & 0 & 6 & 6 \\ 4 & 9 & 4 & -2 & 0 & 1 \\ 4 & 4 & 3 & 0 & 1 & 2 \\ 0 & -2 & 0 & 3 & 2 & 1 \\ 6 & 0 & 1 & 2 & 5 & 0 \\ 6 & 1 & 2 & 1 & 0 & 6 \end{pmatrix},$$

$$\Omega_{17} := \begin{pmatrix} 25 & -2 & 4 & 8 & -1 & 9 \\ -2 & 2 & -2 & 0 & 2 & 0 \\ 4 & -2 & 7 & 3 & -6 & 5 \\ 8 & 0 & 3 & 7 & -1 & 7 \\ -1 & 2 & -6 & -1 & 10 & -5 \\ 9 & 0 & 5 & 7 & -5 & 13 \end{pmatrix}, \ \Omega_{18} := \begin{pmatrix} 5 & 0 & -7 & 3 & 3 & -3 \\ 0 & 10 & 7 & -1 & -8 & -3 \\ -7 & 7 & 28 & 4 & -10 & 0 \\ 3 & -1 & 4 & 25 & 12 & 13 \\ 3 & -8 & -10 & 12 & 18 & 8 \\ -3 & -3 & 0 & 13 & 8 & 22 \end{pmatrix},$$

$$\Omega_{19} := \begin{pmatrix} 18 & -7 & 8 & 7 & 1 & 0 \\ -7 & 14 & -8 & -8 & 6 & -3 \\ 8 & -8 & 14 & 7 & -2 & 10 \\ 7 & -8 & 7 & 15 & 2 & 8 \\ 1 & 6 & -2 & 2 & 6 & 2 \\ 0 & -3 & 10 & 8 & 2 & 24 \end{pmatrix}, \ \Omega_{20} := \begin{pmatrix} 7 & 0 & -2 & -3 & 3 & 2 \\ 0 & 15 & -2 & 4 & -3 & 3 \\ -2 & -2 & 12 & 4 & -4 & -3 \\ -3 & 4 & 4 & 10 & -5 & -6 \\ 3 & -3 & -4 & -5 & 5 & 0 \\ 2 & 3 & -3 & -6 & 0 & 11 \end{pmatrix}.$$

8. References

[1] **Bates, J.M., Granger, C.W.J.** (1969): The combination of forecasts, *Operational Research Quarterly 20*, 451-468.

[2] **Dickinson, J.P.** (1973): Some statistical results in the combination of forecasts, *Operational Research Quarterly 24*, 253-260.

[3] **Klapper, M., Wenzel, T.** (1998): The M-competition forecast data, *Technical Report 21/1998, University of Dortmund*.

[4] **Makridakis et al.** (1982): The accuracy of extrapolation (time series) methods: results of a forecasting competition, *Journal of Forecasting 1*, 111-153.

[5] **Wenzel, T.** (1998): Pitman-closeness and the linear combination of multivariate forecasts, *Technical Report 34/1998, University of Dortmund*.