Combination of biased forecasts: Bias correction or bias based weights?

Thomas Wenzel

Department of Statistics, University of Dortmund, Vogelpothsweg 87

D-44227 Dortmund, Germany

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

Abstract: Most of the literature on combination of forecasts deals with the assumption of unbiased individual forecasts. Here, we consider the case of biased forecasts and discuss two different combination techniques resulting in an unbiased forecast. On the one hand we correct the individual forecasts, and on the other we calculate bias based weights. A simulation study gives some insight in the situations where we should use the different methods.

Key words: combination of forecasts, bias correction, regression, generalized Jackknife, multivariate forecasts.

Acknowledgement: Financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

AMS 1991 Subject classification: 62F10

1. Introduction

The combination of forecasts is usually based on the assumption of unbiased individual forecasts. In the univariate case we restrict the combination weights to sum up to one which results also in an unbiased forecast combination (see e.g. Bates and Granger, 1969). In practice we often have the situation of biased forecasts, which is discussed e.g. in Ehrbeck and Waldmann (1996). If the individual forecasts are biased it is possible to correct them so that we can use the methods for the combination of unbiased forecasts. Another approach is to calculate the combination weights with respect to the bias of the individual forecasts. Here, on the one hand, we can derive the weights considering the covariance matrix as for the MSE-optimal method, and on the other hand, we only use the bias of the individual forecasts. The errors in estimation of the unknown parameters could influence the accuracy of the methods. To analyse this, we perform a simulation study for different situations (different sizes of the bias, stable and unstable covariance structure). Furthermore, we describe the problem for the multivariate case. In this case it is possible to calculate the matrix-mean-square-error optimal unbiased forecast combination which uses the complete covariance structure. Again, we propose bias based combination strategies.

2. Combination of biased forecasts

2.1. The univariate case

We consider the following problem. Let $F_{1,T+1},...,F_{n,T+1}$ be forecasts for a variable Y_{T+1} (T+1: time index) and $u_{i,T+1} := Y_{T+1} - F_{i,T+1},\ i = 1,...,n$ the corresponding forecast errors where $E(u_{i,T+1}) =: \mu_i$ and $Cov(\mathbf{u}_{T+1}) =: \Sigma$, $\mathbf{u}_{T+1} := \left(u_{1.T+1},...,u_{n,T+1}\right)'$. The question is how to combine these possibly biased forecasts to obtain an unbiased forecast. An easy way is a bias correction of the forecasts, that is using $F^*_{i,T+1} := F_{i,T+1} + \mu_i$ and $u^*_{i,T+1} := Y_{T+1} - F^*_{i,T+1}$ which results in $E(u^*_{i,T+1}) = 0,\ i = 1,...,n$. Then it is possible to use weights summing up to one to obtain an unbiased combination, e.g. the simple average of the $F^*_{i,T+1}$'s. The MSE-optimal unbiased combination of the bias corrected forecasts is given by $F^*_{MSE-opt,T+1} := \mathbf{g}_{opt}{}' \mathbf{F}^*_{[T+1]}$, where $\mathbf{g}_{opt} := \left[\left(\mathbf{1}_n{}' \Sigma^{-1} \mathbf{1}_n\right)^{-1} \Sigma^{-1} \mathbf{1}_n\right]$, $\mathbf{F}^*_{[T+1]} := \left(F^*_{1,T+1},...,F^*_{n,T+1}\right)'$ and $\mathbf{1}_n := (1,...,1)'$ is the vector of ones of length n and Σ and μ_i, $i = 1,...,n$, are unknown. Therefore, in practice we have to estimate the bias terms and the covariance matrix for the calculation of this forecast combination. Let $\mathbf{Y} := (Y_1,...,Y_T)'$ be the observed values in the past, $\mathbf{F}_1 := (F_{1,1},...,F_{1,T})',...,\mathbf{F}_n := (F_{n,1},...,F_{n,T})'$ the corresponding forecasts and $\mathbf{u}_i := \mathbf{Y} - \mathbf{F}_i, i = 1,...,n.$ Then we use $\hat{\mu}_i := \bar{\mathbf{u}}_i, i = 1,...,n$ and $\hat{\Sigma} := (\hat{\sigma}_{ij})_{i,j=1,...,n}$ where $\hat{\sigma}_{ij} := \frac{1}{T}\mathbf{u}_i{}'\mathbf{u}_j$, $i, j = 1,...,n$. The calculation of the weights could also be performed on the basis of $\mathbf{F}^* := (\mathbf{F}_1^*,...,\mathbf{F}_n^*)$, where $\mathbf{F}_i^* := \mathbf{F}_i + \hat{\mu}_i \mathbf{1}_T, i = 1,...,n$ and $\mathbf{Y}$. Consulting Granger and Ramanathan (1984) we get:

$$\hat{F}^*_{MSE-opt,T+1} := \left[\left(\mathbf{F}^{*\prime}\mathbf{F}^*\right)^{-1}\mathbf{F}^{*\prime}\mathbf{Y} - \psi\left(\mathbf{F}^{*\prime}\mathbf{F}^*\right)^{-1}\mathbf{1}_n\right]'\tilde{\mathbf{F}}^*_{[T+1]} \quad ,$$

where

$$\psi := \left(\mathbf{1}_n{}'\left(\mathbf{F}^{*\prime}\mathbf{F}^*\right)\mathbf{F}^{*\prime}\mathbf{Y} - 1\right) \Big/ \left(\mathbf{1}_n{}'\left(\mathbf{F}^{*\prime}\mathbf{F}^*\right)^{-1}\mathbf{1}_n\right) \text{ and } \tilde{\mathbf{F}}^*_{[T+1]} := \left(F_{1,T+1} + \hat{\mu}_1,...,F_{n,T+1} + \hat{\mu}_n\right)'.$$

Another approach is to give up the bias correction. In the following we use the bias directly to calculate the combination weights. We present a technique which is also based on the covariance matrix and other techniques which disregard the covariance structure and use only the bias terms.

Theorem 1: Let $F_{1,T+1},...,F_{n,T+1}$ be forecasts for Y_{T+1} and $u_{i,T+1} := Y_{T+1} - F_{i,T+1}$ be the individual forecast errors where $E(u_{i,T+1}) =: \mu_i$, $i = 1,...,n$. Further, let $Cov(\mathbf{u}_{T+1}) =: \Sigma$, Σ p.d., where $\mathbf{u}_{T+1} := (u_{1,T+1},...,u_{n,T+1})'$. We assume that there exists at least one (i,j), $i, j \in \{1,...,n\}$, $i < j$, where $\mu_i \neq \mu_j$. The MSE-optimal unbiased forecast combination of the form $F_{w,T+1} := \mathbf{w}'\mathbf{F}_{[T+1]}$ where $\mathbf{w}'\mathbf{1}_n = 1$ and $\mathbf{F}_{[T+1]} := (F_{1,T+1},...,F_{n,T+1})'$ is given by

$$\mathbf{w}_{opt} := \frac{\left(\mathbf{1}_n{}'\Sigma^{-1}\mu\right)\Sigma^{-1}\mu - \left(\mu'\Sigma^{-1}\mu\right)\Sigma^{-1}\mathbf{1}_n}{\left(\mathbf{1}_n{}'\Sigma^{-1}\mu\right)^2 - \left(\mu'\Sigma^{-1}\mu\right)\left(\mathbf{1}_n{}'\Sigma^{-1}\mathbf{1}_n\right)}, \text{ where } \mu := (\mu_1,...,\mu_n)'.$$

Proof: We restrict the weights to sum up to one, which means $\mathbf{w}'\mathbf{1}_n = 1$. We also want to minimize the MSE subject to the requirement of an unbiased combination which can be expressed in $\mathbf{w}'\mu = 0$. Thus, we consider the following function:

$$L(w,\lambda,\varphi) := \mathbf{w}'\Sigma\mathbf{w} - \lambda(\mathbf{w}'\mu) - \varphi(\mathbf{w}'\mathbf{1}_n - 1).$$

The necessary conditions for a minimum are

I) $$\frac{\delta L(\mathbf{w},\lambda,\varphi)}{\delta \mathbf{w}} = 2\mathbf{w}'\Sigma - \lambda\mu' - \varphi\mathbf{1}_n{}' \overset{!}{=} \mathbf{0}'$$

II) $$\frac{\delta L(\mathbf{w},\lambda,\varphi)}{\delta \lambda} = -\mathbf{w}'\mu \overset{!}{=} 0$$

III) $$\frac{\delta L(\mathbf{w},\lambda,\varphi)}{\delta \varphi} = 1 - \mathbf{w}'\mathbf{1}_n \overset{!}{=} 0$$

From I) we get

$$\mathbf{w}' = \frac{\lambda}{2}\mu'\Sigma^{-1} + \frac{\varphi}{2}\mathbf{1}_n{}'\Sigma^{-1}$$

and inserting in II) and III) gives

$$\frac{\lambda}{2}\mu'\Sigma^{-1}\mu + \frac{\varphi}{2}\mathbf{1}_n{}'\Sigma^{-1}\mu = 0 \qquad \text{and}$$

$$\frac{\lambda}{2}\mu'\Sigma^{-1}\mathbf{1}_n + \frac{\varphi}{2}\mathbf{1}_n{}'\Sigma^{-1}\mathbf{1}_n = 1.$$

Some easy calculations result in

$$\frac{\lambda}{2} = \frac{\mathbf{1}_n{}'\Sigma^{-1}\mu}{\left(\mathbf{1}_n{}'\Sigma^{-1}\mu\right)^2 - \left(\mu'\Sigma^{-1}\mu\right)\left(\mathbf{1}_n{}'\Sigma^{-1}\mathbf{1}_n\right)} \quad \text{and}$$

$$\frac{\varphi}{2} = \frac{-\left(\mu'\Sigma^{-1}\mu\right)}{\left(\mathbf{1}_n{}'\Sigma^{-1}\mu\right)^2 - \left(\mu'\Sigma^{-1}\mu\right)\left(\mathbf{1}_n{}'\Sigma^{-1}\mathbf{1}_n\right)} \; .$$

Because of the Cauchy-Schwarz inequality the denominator of the preceding expressions is non-positive. Since the μ_i's are not all equal and therefore μ and $\mathbf{1}_n$ are linearly independent it is even negative.

Refering to I) the optimal weight vector turns out to be

$$\mathbf{w}_{opt} := \frac{\left(\mathbf{1}_n{}'\Sigma^{-1}\mu\right)\Sigma^{-1}\mu - \left(\mu'\Sigma^{-1}\mu\right)\Sigma^{-1}\mathbf{1}_n}{\left(\mathbf{1}_n{}'\Sigma^{-1}\mu\right)^2 - \left(\mu'\Sigma^{-1}\mu\right)\left(\mathbf{1}_n{}'\Sigma^{-1}\mathbf{1}_n\right)}.$$

Looking at the form of the function $L(\mathbf{w},\lambda,\varphi)$ it is straightforward that $\mathbf{w}_{opt}$ is the minimizing vector.

In practice we have to calculate $\hat{F}_{w_{opt},T+1}$ by using the estimators $\hat{\Sigma}$ and $\hat{\mu}$ as above. Here it is also possible to estimate the weights directly by restricted regression. Using $\mathbf{R} := \begin{pmatrix} \mathbf{1}_T{}'\mathbf{F} \\ \mathbf{1}_n{}' \end{pmatrix}$, where $\mathbf{F} := (\mathbf{F}_1,...,\mathbf{F}_n)$ and $\mathbf{r} := \begin{pmatrix} \mathbf{1}_T{}'\mathbf{Y} \\ 1 \end{pmatrix}$ the optimal weights are given by

$$\mathbf{w}_{opt} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{Y} - (\mathbf{F}'\mathbf{F})^{-1}\mathbf{R}'\left(\mathbf{R}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{R}'\right)^{-1}\left(\mathbf{R}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{Y} - \mathbf{r}\right).$$

Of course, the combination $F_{w_{opt},T+1}$ has not a smaller MSE than the combination $F^{*}_{MSE-opt,T+1}$. Since both are unbiased forecasts their MSEs are given by the error variances $\mathbf{w}_{opt}{}'\Sigma\mathbf{w}_{opt}$ and $\mathbf{g}_{opt}{}'\Sigma\mathbf{g}_{opt}$, respectively, and the vector $\mathbf{w}_{opt}$ includes one more restriction than the vector $\mathbf{g}_{opt}$. But we have to remark that in practice it might be difficult to justify a bias correction. In this situation we correct forecasts given by some experts or calculated by sophisticated and expensive models before combining them and thus we have to convince the analyst that he cannot use the individual forecasts as they are.

In the following we present bias based methods which disregard the covariance structure and also result in an unbiased combination.

<u>Theorem 2:</u> Let $F_{1,T+1},...,F_{n,T+1}$ be forecasts for Y_{T+1} and $u_{i,T+1} := Y_{T+1} - F_{i,T+1}$ be the individual forecast errors where $E(u_{i,T+1}) =: \mu_i$, $i = 1,...,n$. Further, let $\mu_j \neq 0$ and $\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq v}}^{n} R_j \neq 1$, where

$R_j := \frac{\mu_v}{\mu_j}$, $j = 1,...,n,\ j \neq v,\ v \in \{1,...,n\}$ fixed but arbitrary. Then

$$F_{J1,T+1} := \frac{F_{v,T+1} - \frac{1}{n-1}\sum_{\substack{j=1\\ j\neq v}}^{n} R_j F_{j,T+1}}{1 - \frac{1}{n-1}\sum_{\substack{j=1\\ j\neq v}}^{n} R_j}$$

is an unbiased forecast for Y_{T+1}.

Proof: For the error of the forecast combination we get

$$u_{J1,T+1} := Y_{T+1} - F_{J1,T+1} = \frac{1}{1 - \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq v}}^{n} R_j}\left[\left(1 - \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq v}}^{n} R_j\right)Y_{T+1} - F_{v,T+1} + \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq v}}^{n} R_j F_{j,T+1}\right]$$

$$= \frac{1}{1 - \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq v}}^{n} R_j}\left[Y_{T+1} - F_{v,T+1} - \left(\frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq v}}^{n} R_j (Y_{T+1} - F_{j,T+1})\right)\right]$$

$$= \frac{1}{1 - \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq v}}^{n} R_j}\left[u_{v,T+1} - \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq v}}^{n} R_j u_{j,T+1}\right] \quad .$$

Thus

$$E(u_{J1,T+1}) = \frac{1}{1 - \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq v}}^{n} R_j}\left[E(u_{v,T+1}) - \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq v}}^{n} \frac{\mu_v}{\mu_j} E(u_{j,T+1})\right]$$

$$= \frac{1}{1 - \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq v}}^{n} R_j}\left[\mu_v - \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq v}}^{n} \frac{\mu_v}{\mu_j}\mu_j\right]$$

$$= 0 \quad .$$

The bias based method is in the form of the generalized Jackknife-estimator well-known in point estimation. Its MSE is equal or exceeds the MSE of the MSE-optimal combination, but in practice one has to estimate the unknown parameters for the calculation of the combined forecast. The errors in estimation could result in more unreliability of the MSE-optimal forecast combination because it depends also on the whole covariance structure. We have to remark that all of the methods presented above might result in negative weights and produce extreme outliers. An example for this, regarding e.g. the MSE-optimal combination of unbiased forecasts, is given in Klapper (1998). Hence, there is the demand of a more robust bias based combination strategy.

Theorem 3: Let $F_{1,T+1},\ldots,F_{n,T+1}$ be forecasts for Y_{T+1} and $u_{i,T+1} := Y_{T+1} - F_{i,T+1}$, where $E(u_{i,T+1}) =: \mu_i$, $i = 1,\ldots,n$. Further, let $\mu_j \neq 0$ and $\sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j| \neq -\sum_{\substack{j=1 \\ j\neq v}}^{n} \text{sign}(R_j)$, where $R_j := \frac{\mu_v}{\mu_j}$, $j = 1,\ldots,n,\ j \neq v,\ v \in \{1,\ldots,n\}$ fixed but arbitrary. Then:

i) $F_{J2,T+1} := \dfrac{\gamma F_{v,T+1} + \sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j| F_{j,T+1}}{\gamma + \sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j|}$, where $\gamma := -\sum_{\substack{j=1 \\ j\neq v}}^{n} \text{sign}(R_j)$, is an unbiased forecast for Y_{T+1}.

ii) If there exists at least one $\mu_i > 0$ and at least one $\mu_j < 0$, $i \neq j,\ i, j \in \{1,\ldots,n\}$, then we construct $F_{J2,T+1}$ as follows as an unbiased forecast with value inside the interval of the individual forecasts:

If there exists an unbiased individual forecast $F_{k,T+1}$, $k \in \{1,\ldots,n\}$ by definition $v := k$, else

if $\#(\mu_i > 0)_{i=1,\ldots,n} \overset{(>)}{\geq} \frac{n}{2}$ then choose v so that $\mu_v < 0$, else

if $\#(\mu_i > 0)_{i=1,\ldots,n} \overset{(\leq)}{<} \frac{n}{2}$ then choose v so that $\mu_v > 0$.

Proof:

i) At first we calculate the error of the forecast combination, and then we show that it has mean zero.

$$u_{J2,T+1} := Y_{T+1} - F_{J2,T+1} = \frac{1}{\gamma + \sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j|} \left[\left(\gamma + \sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j| \right) Y_{T+1} - \gamma F_{v,T+1} - \sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j| F_{j,T+1} \right]$$

$$= \frac{1}{\gamma + \sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j|} \left[\gamma \left(Y_{T+1} - F_{v,T+1} \right) + \sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j| \left(Y_{T+1} - F_{j,T+1} \right) \right]$$

$$= \frac{1}{\gamma + \sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j|} \left[\gamma u_{v,T+1} + \sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j| u_{j,T+1} \right] \quad \text{and thus}$$

$$E(u_{J2,T+1}) = \frac{1}{\gamma + \sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j|} \left[\gamma E(u_{v,T+1}) + \sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j| E(u_{j,T+1}) \right]$$

$$= \frac{1}{\gamma + \sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j|} \left[-\sum_{\substack{j=1 \\ j\neq v}}^{n} \operatorname{sign}(R_j)\mu_v + \sum_{\substack{j=1 \\ j\neq v}}^{n} \left| \frac{\mu_v}{\mu_j} \right| \mu_j \right]$$

$$= \frac{1}{\gamma + \sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j|} \left[-\sum_{\substack{j=1 \\ j\neq v}}^{n} \operatorname{sign}(R_j)\mu_v + \sum_{\substack{j=1 \\ j\neq v}}^{n} \operatorname{sign}(R_j) \frac{\mu_v}{\mu_j} \mu_j \right] = 0 \quad .$$

ii) The special choice of $F_{v,T+1}$ (respectively μ_v) guarantees that $\gamma \geq 0$, since in the case where none of the forecasts is unbiased, the number of μ_j's with different sign as μ_v is greater or equal than the number of μ_j's with the same sign as μ_v. Therefore, by definition all weights are in the interval [0,1] and sum up to one.

<u>Remark:</u> For the cases where all $\mu_j>0$ or all $\mu_j<0$ we get $\gamma = -(n-1)$. This does not depend on the choice of $F_{v,T+1}$. If we choose v so that $|\mu_v| = \max_{i=1,\dots,n} |\mu_i|$ we get $\sum_{\substack{j=1 \\ j\neq v}}^{n} |R_j| = \sum_{\substack{j=1 \\ j\neq v}}^{n} \left| \frac{\mu_v}{\mu_j} \right| > n-1$

and hence only the weight for $F_{v,T+1}$ is negative.

2.2. Simulation study

We consider the combination of six biased forecasts by using two different bias vectors: $b_1 := (50,40,20,10,-10,-20)'$ and $b_2 := (5,4,2,1,-1,-2)'$. Furthermore, we randomly generate 20 covariance matrices and on their basis (together with the bias) 200 series (6 forecasts) of normally distributed forecast errors are generated. The series are of length 60. We fix 10 data points to calculate the first combination weights, thus 50 performance points are left for our analysis. In each step we calculate the new weights by regarding all available history for the estimation of the unknown parameters. To compare the different methods we calculate their RMSEs relative to the values of the simple average of the individual forecasts. The study includes the following methods: 6 bias corrected individual forecasts (No. 1-6), MSE-optimal combination with the assumption of unbiased individual forecasts (No. 7), MSE-optimal combination $F^*_{MSE-opt,T+1}$ of bias corrected forecasts (No. 8), MSE-optimal combination $F_{w_{opt},T+1}$ of the biased individual forecasts (No. 9), simple average (No. 10), simple average of bias corrected forecasts (No. 11), and the two bias based combinations $F_{J1,T+1}$ (No. 12) and $F_{J2,T+1}$ (No. 13). For the combination $F_{J1,T+1}$ we choose the individual forecast with the smallest absolute bias as $F_{v,T+1}$, and in addition for the combination $F_{J2,T+1}$ we choose the candidate with the highest absolute bias as $F_{v,T+1}$. Instead of calculating all data points with stable covariance matrices we consider a situation of structural change. Here, the variances of the individual forecast errors are varying over time which is described in detail below.

a) time stable covariance structure

a1) bias vector b_1

<u>Table 1:</u> Comparison of methods for case a1

S1: number of times simple average is beaten, S2: number of times simple average of bias corrected forecasts is beaten, best: number of times the special method is the best one. M1,...,M13 denote the methods.

Cov. No.		M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
1	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	200	200	200	0	-	0	151
	best	0	0	0	0	0	0	4	196	0	0	0	0	0
2	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	11	200	200	200	0	-	3	116
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
3	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	76	132	74	0	-	3	82
	best	0	0	0	0	0	0	12	109	4	0	43	0	32
4	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	200	200	200	0	-	0	168
	best	0	0	0	0	0	0	4	193	3	0	0	0	0

Table 1 continiued														
5	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	199	199	199	0	-	0	0
	best	0	0	0	0	0	0	4	193	3	0	0	0	0
6	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	7	0	0	0	18	200	18	0	-	0	1
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
7	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	103	200	103	0	-	0	0
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
8	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	3	0	6	0	0	198	192	198	0	-	0	0
	best	0	0	0	0	0	0	74	72	52	0	2	0	0
9	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	129	98	29	0	-	0	93
	best	0	0	0	0	0	0	36	27	52	0	45	0	40
10	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	132	187	133	0	-	0	0
	best	0	0	0	0	0	0	3	184	0	0	13	0	0
11	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	200	200	200	0	-	0	0
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
12	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	0	126	0	0	-	0	0
	best	0	0	0	0	0	0	0	126	0	0	74	0	0
13	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	191	200	191	0	-	0	3
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
14	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	3	0	200	200	200	0	-	0	156
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
15	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	1	0	200	200	200	0	-	0	190
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
16	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	199	199	199	0	-	0	0
	best	0	0	0	0	0	0	67	77	56	0	0	0	0
17	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	200	200	200	0	-	0	0
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
18	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	0	200	0	0	-	0	0
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
19	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	39	0	160	178	160	0	-	22	124
	best	0	0	0	0	2	0	10	134	23	0	13	0	18
20	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	200	200	200	0	-	0	107
	best	0	0	0	0	0	0	1	196	3	0	0	0	0

If we consider Table 1 it is obvious that method No. 8 is best in the sense of the RMSE. In 18 cases it is the best one. For covariance matrix No. 8 methods No. 7, 8 and 9 and for covariance matrix No. 9 methods No. 9, 11, 13, 7 and 8 perform similarly. This result is not a surprise because of the time stable covariance structure. With this assumption, method No. 14 is theoretically optimal and the estimators for the unknown parameters perform well. If we compare method No. 13 ($F_{J2,T+1}$) and method No. 11 (simple average of bias corrected forecasts) we can see that for covariance matrices No. 1, 2, 4, 14, 15, 19 and 20 the first one performs better. These are exactly the cases (also covariance matrix No. 9) where the bias based combination theoretically outperforms the simple average of bias corrected forecasts.

We can also see that neglecting the bias and the covariance structure, the simple average combination is of less quality.

a2) bias vector b_2

Table 2: Comparison of methods for case a2

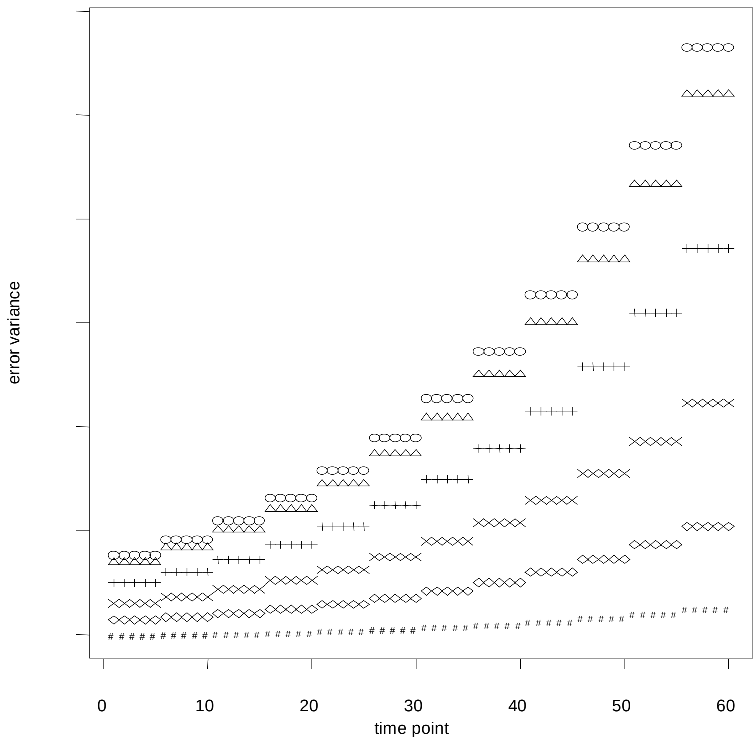
Cov. No.		M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
1	S1	0	0	6	0	0	0	200	200	200	-	200	0	164
	S2	0	0	0	0	0	0	200	200	200	0	-	0	20
	best	0	0	0	0	0	0	5	194	1	0	0	0	0
2	S1	0	0	0	3	0	117	200	200	200	-	199	45	163
	S2	0	0	0	0	0	18	200	200	200	1	-	5	92
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
3	S1	0	0	0	3	0	23	193	198	191	-	200	22	171
	S2	0	0	0	0	0	0	78	131	75	0	-	0	40
	best	0	0	0	0	0	0	11	109	9	0	57	0	14
4	S1	0	0	138	0	0	0	200	200	200	-	200	0	138
	S2	0	0	0	0	0	0	200	200	200	0	-	0	12
	best	0	0	0	0	0	0	8	190	2	0	0	0	0
5	S1	0	24	37	0	0	6	200	200	200	-	200	11	133
	S2	0	0	0	0	0	0	200	200	200	0	-	0	0
	best	0	0	0	0	0	0	6	191	3	0	0	0	0
6	S1	0	0	200	193	0	0	200	200	200	-	200	0	172
	S2	0	0	11	0	0	0	26	200	20	0	-	0	1
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
7	S1	0	0	1	14	0	0	200	200	200	-	200	0	65
	S2	0	0	0	0	0	0	95	200	87	0	-	0	0
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
8	S1	0	184	0	166	1	0	200	200	200	-	200	1	133
	S2	0	1	0	7	0	0	198	194	198	0	-	0	1
	best	0	0	0	0	0	0	60	64	74	0	2	0	0
9	S1	21	38	21	0	6	11	200	200	200	-	200	22	175
	S2	0	0	0	0	0	0	130	113	130	0	-	0	54
	best	0	0	0	0	0	0	35	37	44	0	50	0	34
10	S1	1	0	0	0	0	0	200	200	200	-	200	0	91
	S2	0	0	0	0	0	0	146	185	130	0	-	0	0
	best	0	0	0	0	0	0	4	178	4	0	14	0	0
11	S1	0	79	1	0	0	0	200	200	200	-	200	0	44
	S2	0	0	0	0	0	0	200	200	200	0	-	0	0
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
12	S1	0	0	0	0	7	7	200	200	200	-	200	35	198
	S2	0	0	0	0	0	0	0	144	0	0	-	0	0
	best	0	0	0	0	0	0	0	144	0	0	56	0	0
13	S1	0	0	5	0	0	0	200	200	200	-	200	1	177
	S2	0	0	0	0	0	0	199	200	193	0	-	0	4
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
14	S1	0	0	5	0	126	46	200	200	200	-	200	58	188
	S2	0	0	0	0	1	0	200	200	200	0	-	0	106
	best	0	0	0	0	0	0	2	198	0	0	0	0	0
15	S1	6	0	2	0	35	52	200	200	200	-	200	44	168
	S2	0	0	0	0	1	0	200	200	200	0	-	2	95
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
16	S1	0	0	0	32	44	0	200	200	200	-	200	19	185
	S2	0	0	0	0	0	0	199	200	198	0	-	0	1
	best	0	0	0	0	0	0	65	69	66	0	0	0	0
17	S1	0	7	35	0	48	0	200	200	200	-	200	2	192
	S2	0	0	0	0	0	0	200	200	200	0	-	0	0
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
18	S1	0	52	0	0	0	0	200	200	188	-	200	0	8
	S2	0	0	0	0	0	0	0	200	0	0	-	0	0
	best	0	0	0	0	0	0	0	200	0	0	0	0	0
19	S1	4	0	0	0	192	0	200	200	200	-	200	97	196
	S2	0	0	0	0	42	0	173	187	168	0	-	12	86
	best	0	0	0	0	0	0	23	132	31	0	6	0	8
20	S1	0	0	11	0	0	18	200	200	200	-	200	4	179
	S2	0	0	0	0	0	0	200	200	200	0	-	0	29
	best	0	0	0	0	0	0	2	196	2	0	0	0	0

Although for the same covariance matrices as above the combination $F_{J2,T+1}$ should be better than the simple average of the bias corrected forecasts, it only happens in case No. 14. In some of these cases it is clearly outperformed. Naturally, the best combination is again method No. 8. In cases No. 8, 9 and 16 some methods are nearly of the same high quality (methods No. 9, 8, 7, methods No. 11, 9, 8, 7, 13 and methods No. 8, 9, 7). Because of the "low" bias the simple average performs better than before, whereas method No. 12 ($F_{J1,T+1}$) is again of poor quality.

b) Structural change all five data points

We analyse a structural change every five steps. We generate first five data points by using $\Sigma_{(1)} := \Sigma$ as before. We generate the next five points with $\Sigma_{(2)} := \Sigma_{(1)} + 0.2 \cdot \mathrm{diag}\left(\Sigma_{(1)}\right)$ where $\mathrm{diag}\left(\Sigma_{(1)}\right)$ is a diagonal matrix of the diagonal elements of $\Sigma_{(1)}$. Then we calculate five points with $\Sigma_{(3)} := \Sigma_{(2)} + 0.2 \cdot \mathrm{diag}\left(\Sigma_{(2)}\right)$, and so on. Thus, only the variances will change over time which is illustrated in Figure 1. The differences between the error variances increase, so over time the quality of all forecasts decreases but the forecasts with lower variance are less influenced by the changes.

Figure 1: Structural changes in the error variances



b1) bias vector b_1

Table 3: Comparison of methods for case b1

Cov. No.		M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
1	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	9	83	9	0	-	0	14
	best	0	0	0	0	0	0	0	79	0	0	114	0	7
2	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	2	68	76	68	0	-	1	15
	best	0	0	0	0	0	0	28	37	19	0	110	0	6
3	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	10	9	10	0	-	0	36
	best	0	0	0	0	0	0	2	1	1	0	161	0	35
4	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	7	0	0	0	199	199	199	0	-	0	52
	best	0	0	0	0	0	0	55	103	41	0	1	0	0
5	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	0	103	0	0	-	0	0
	best	0	0	0	0	0	0	0	103	0	0	97	0	0
6	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	49	6	0	0	0	195	1	0	-	0	0
	best	0	0	1	0	0	0	0	194	0	0	5	0	0
7	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	0	14	0	0	-	0	0
	best	0	0	0	0	0	0	0	14	0	0	186	0	0
8	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	5	0	6	0	0	34	166	33	0	-	0	8
	best	0	1	0	2	0	0	1	161	0	0	33	0	2
9	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	71	80	71	0	-	0	97
	best	0	0	0	0	0	0	26	32	17	0	70	0	55
10	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	37	45	35	0	0	0	57
	best	0	0	0	0	0	0	5	23	8	0	119	0	45
11	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	0	137	0	0	-	0	0
	best	0	0	0	0	0	0	0	137	0	0	63	0	0
12	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	24	52	23	0	-	0	20
	best	0	0	0	0	0	0	5	43	0	0	139	0	13
13	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	2	91	1	0	-	0	0
	best	0	0	0	0	0	0	0	91	0	0	109	0	0
14	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	92	146	92	0	-	0	74
	best	0	0	0	0	0	0	7	129	5	0	39	0	20
15	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	121	125	123	0	-	0	158
	best	0	0	0	0	0	0	18	35	19	0	28	0	100
16	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	1	0	0	22	154	22	0	-	0	0
	best	0	0	0	0	0	0	0	154	0	0	46	0	0
17	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	22	114	22	0	.	0	5
	best	0	0	0	0	0	0	1	112	1	0	85	0	1
18	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	3	0	0	0	0	0	158	0	0	-	0	0
	best	0	0	0	0	0	0	0	158	0	0	42	0	0
19	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	14	0	89	99	89	0	-	4	127
	best	0	0	0	0	3	0	15	48	11	0	47	0	76
20	S1	0	200	200	200	200	200	200	200	200	-	200	200	200
	S2	0	0	0	0	0	0	91	96	90	0	-	0	85
	best	0	0	0	0	0	0	20	41	12	0	77	0	50

We can see that the structural change in the variances has influence on methods No. 7, No. 8 and No. 9. Because these strategies depend on the covariance structure, the errors in estimation occuring in this case worsen their quality. Now the bias based method No. 13 in two cases is better than the arithmetic mean of bias corrected forecasts. As a result of the special structural change it should now outperform method No. 11 only in the cases No. 9, 15 and 19. For instance, method No. 11 is nine times, the MSE-optimal combination of bias corrected forecasts is eight times, and the combination $F_{J2,T+1}$ is two times the best (covariance matrices No. 15 and 19). In one case the MSE-optimal combination of bias corrected forecasts and the simple average of bias corrected forecasts are best. For the cases No. 9, No. 19 and No. 20 the differences between the best and some other methods are smaller.

b2) bias vector b_2

Table 4: Comparison of methods for case b2

Cov. No.		M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
1	S1	0	0	0	0	0	0	117	179	72	-	198	0	23
	S2	0	0	0	0	0	0	23	73	12	2	-	0	2
	best	0	0	0	0	0	0	4	70	1	1	123	0	1
2	S1	0	0	0	0	0	3	164	169	146	-	194	1	51
	S2	0	0	0	0	0	1	69	74	60	6	-	0	14
	best	0	0	0	0	0	0	30	28	23	4	108	0	7
3	S1	0	0	0	0	0	0	90	102	68	-	195	0	57
	S2	0	0	0	0	0	0	13	10	11	5	-	0	6
	best	0	0	0	0	0	0	8	5	5	5	173	0	4
4	S1	39	0	52	0	0	0	200	200	200	-	200	0	20
	S2	4	0	8	0	0	0	199	199	100	0	-	0	1
	best	0	0	0	0	0	0	66	96	37	0	1	0	0
5	S1	0	0	0	0	0	0	92	198	43	-	200	0	10
	S2	0	0	0	0	0	0	2	114	1	0	-	0	0
	best	0	0	0	0	0	0	0	114	0	0	86	0	0
6	S1	0	0	143	61	0	0	105	200	24	-	200	0	3
	S2	0	0	48	7	0	0	2	195	0	0	-	0	0
	best	0	0	0	0	0	0	0	195	0	0	5	0	0
7	S1	0	0	0	0	0	0	23	111	11	-	195	0	4
	S2	0	0	0	0	0	0	1	14	0	5	-	0	1
	best	0	0	0	0	0	0	1	13	0	4	181	0	1
8	S1	0	38	0	23	0	0	175	199	137	-	200	0	21
	S2	0	10	0	5	0	0	62	176	37	0	-	0	2
	best	0	0	0	0	0	0	2	174	0	0	24	0	0
9	S1	2	0	0	0	0	0	186	159	174	-	197	0	78
	S2	0	0	0	0	0	0	102	70	93	3	-	0	26
	best	0	0	0	0	0	0	56	20	28	0	86	0	10
10	S1	1	0	0	0	0	0	130	142	106	-	197	0	30
	S2	0	0	0	0	0	0	39	42	32	3	-	0	5
	best	0	0	0	0	0	0	16	23	14	0	146	0	1
11	S1	0	11	0	0	0	0	15	198	5	-	200	0	0
	S2	0	0	0	0	0	0	0	136	0	0	-	0	0
	best	0	0	0	0	0	0	0	136	0	0	64	0	0
12	S1	0	0	0	0	0	0	175	196	141	-	200	1	103
	S2	0	0	0	0	0	0	31	57	23	0	-	0	6
	best	0	0	0	0	0	0	7	46	4	0	139	0	4
13	S1	0	0	0	0	0	0	92	190	36	-	199	0	10
	S2	0	0	0	0	0	0	9	85	4	1	-	0	0
	best	0	0	0	0	0	0	1	84	0	0	115	0	0
14	S1	0	0	0	0	14	1	198	197	182	-	199	3	120
	S2	0	0	0	0	2	0	126	170	103	1	-	0	38
	best	0	0	0	0	0	0	23	135	9	0	25	0	8
15	S1	0	0	0	0	0	1	185	187	174	-	194	0	94
	S2	0	0	0	0	0	0	126	127	123	6	-	0	37
	best	0	0	0	0	0	0	54	42	34	1	57	0	12
16	S1	0	0	0	1	2	0	182	199	118	-	200	0	25
	S2	0	0	0	0	1	0	66	161	30	0	-	0	0
	best	0	0	0	0	0	0	2	157	2	0	39	0	0

Table 4 continiued														
17	S1	0	0	1	0	0	0	196	200	190	-	200	0	115
	S2	0	0	0	0	0	0	41	121	33	0	-	0	4
	best	0	0	0	0	0	0	0	120	1	0	78	0	1
18	S1	0	7	0	0	0	0	31	196	6	-	198	0	0
	S2	0	3	0	0	0	0	1	152	0	2	-	0	0
	best	0	0	0	0	0	0	0	152	0	0	48	0	0
19	S1	0	0	0	0	60	0	188	189	180	-	198	17	120
	S2	0	0	0	0	8	0	109	117	94	2	-	1	57
	best	0	0	0	0	0	0	47	48	17	0	67	0	21
20	S1	0	0	0	0	0	1	194	193	184	-	200	1	101
	S2	0	0	0	0	0	0	90	89	83	0	-	0	28
	best	0	0	0	0	0	0	32	35	31	0	92	0	10

Here, the bias based combination techniques are of poor quality. Using the simple average of bias corrected forecasts or the MSE-optimal combination of bias corrected forecasts is more accurate. Method No. 8 in nine cases is the best, method No. 11 in eight cases (adding the following three). In one case (covariance matrix No. 9) methods No. 11 and 7 perform better, for covariance matrix No. 15 methods No. 11, 7, 8, 9 and for covariance matrix No. 19 methods No. 11, 8, 7. Here, the simple average of the individual forecast is of higher quality than in b3.

2.3. Concluding remarks for the univariate case

If the covariance structure is stable over time the MSE-optimal combination is of course the best in the sense of the RMSE. Depending on the covariance structure in the case of "large" absolute bias and so "large" distances between the bias, the combination $F_{J2,T+1}$ can outperform the simple average of bias corrected individual forecasts. When the absolute bias are "small" and so the distances are "small", too, more often the "wrong" individual forecast is chosen as $F_{v,T+1}$. Furthermore, we frequently get a "wrong" γ. Due to the given covariance matrices in this simulation study the combination $F_{J2,T+1}$ performs poorly. If a structural change happens at all five data points in the error variances, the simple average of bias corrected forecasts performs as good as the MSE-optimal combination of bias corrected forecasts. The combination $F_{J2,T+1}$ performs better than the other methods in the situation of "large" bias and where it is, theoretically, of high quality. Furthermore, the given covariance matrices in this simulation study are a reason for the bad performance of the method $F_{J1,T+1}$.

We have to remark that the simulation study is giving only limited insight into the characteristics of the different methods. Other structural changes, e.g. in the covariances between the forecast errors are possible. A more extensive analysis of this problem, regarding other methods, is given e.g. in Diebold and Pauly (1987) or in Deutsch, Granger and Teräsvirta (1994). Nevertheless, if the differences between the bias are not too "small" we can use knowledge from the past to decide if we calculate a combination of bias corrected forecast or a bias based forecast combination. Furthermore, if we consider bias corrected forecasts for a combination, then the question arises if the forecasting models must be respecified. On the other side, in bias based combinations we use the forecast as they are and give them special weights.

Finally, if we look at the combinations $F_{J1,T+1}$ and $F_{J2,T+1}$, we notice that other strategies in the choice of $F_{v,T+1}$ are possible. For this we can again take advantage of experience from the past.

2.4 The multivariate case

Let $\mathbf{Y}_{T+1} := \left(Y_{1,T+1},\dots,Y_{k,T+1}\right)'$, $k \geq 2$, be a vector to be forecasted, $\mathbf{F}_{1,T+1},\dots,\mathbf{F}_{n,T+1}$ be forecasts, where $\mathbf{F}_{i,T+1} := \left(F^{(i)}_{1,T+1},\dots,F^{(i)}_{k,T+1}\right)'$ and $\mathbf{u}_{i,T+1} := \mathbf{Y}_{T+1} - \mathbf{F}_{i,T+1}$, with $E\left(\mathbf{u}_{i,T+1}\right) =: \mu_i$ and $\mu_i := (\mu_{i1},\dots,\mu_{ik})'$, $i = 1,\dots,n$. Further, let $\mathbf{u}_{T+1} := \left(\mathbf{u}_{1,T+1}{}',\dots,\mathbf{u}_{n,T+1}{}'\right)'$ and $\Sigma := (\Sigma_{rs})_{r,s=1,\dots,n} := \mathrm{Cov}\left(\mathbf{u}_{T+1}\right)$. We want to calculate an unbiased forecast combination where we use weight matrices $\mathbf{G}_i \sim (k \times k)$, $i = 1,\dots,n$, summing up to $\mathbf{I}_k$. An easy way, like in the univariate case, is to consider the bias corrected forecasts. Then, the optimal weight matrices minimizing the matrix-mean-square-error (MMSE) of the combined forecast in the sense of the Löwner-ordering are given by (see e.g. Wenzel, 1998)

$$\mathbf{G}_{opt} := [\mathbf{G}_{1,opt},\dots,\mathbf{G}_{n,opt}] := [\mathbf{W}'\mathbf{V}^{-1}, \mathbf{I}_k - \mathbf{W}'\mathbf{V}^{-1}\mathbf{I}^*_k] \ ,$$

where

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

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

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

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

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

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

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

As in the univariate case we now calculate a MMSE-optimal unbiased forecast combination without using a bias correction.

Theorem 4: Let $\mathbf{F}_{i,T+1} := \left(F^{(i)}_{1,T+1},...,F^{(i)}_{k,T+1}\right)'$ be forecasts for $\mathbf{Y}_{T+1} := \left(Y_{1,T+1},...,Y_{k,T+1}\right)'$, $k \geq 2$, and $\mathbf{u}_{i,T+1} := \mathbf{Y}_{T+1} - \mathbf{F}_{i,T+1}$, where $E\left(\mathbf{u}_{i,T+1}\right) =: \mu_i$, $i = 1,...,n$. Further let $\mathbf{u}_{T+1} := \left(\mathbf{u}_{1,T+1}{}',...,\mathbf{u}_{n,T+1}{}'\right)'$ and $\Sigma := (\Sigma_{rs})_{r,s=1,...,n} := \mathrm{Cov}(\mathbf{u}_{T+1})$. Assume that $\gamma = \left((\mu_1 - \mu_n)',...,(\mu_{n-1} - \mu_n)'\right)' \neq \mathbf{0}$. The MMSE-optimal (in the sense of the Löwner-ordering) unbiased forecast combination of the form $\mathbf{F}_{H,T+1} := \sum_{i=1}^{n} \mathbf{H}_i \mathbf{F}_{i,T+1}$, where $\sum_{i=1}^{n} \mathbf{H}_i = \mathbf{I}_k$, is given by

$$\mathbf{H}_{opt} := [\mathbf{H}_{1,opt},...,\mathbf{H}_{n,opt}] := [(\mathbf{W}' + \mathbf{D})\mathbf{V}^{-1}, \mathbf{I}_k - (\mathbf{W}' + \mathbf{D})\mathbf{V}^{-1}\mathbf{I}^*_k],$$

where $\mathbf{D} := \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\left(\mu_n - \mathbf{W}'\mathbf{V}^{-1}\gamma\right)\gamma'$.

Proof: Because the MMSE of the optimal forecast combination must have minimal trace we minimize it in the following and prove afterwards, that for any other combination which satisfies the restrictions, the optimal MMSE-combination has smaller or equal MMSE in the sense of the Löwner-ordering. Consulting Odell et al. (1989), the MMSE of any combination which satisfies the restrictions (1) $\sum_{i=1}^{n} \mathbf{H}_i = \mathbf{I}_k$ and (2) $\sum_{i=1}^{n-1} \mathbf{H}_i(\mu_i - \mu_n) = \mu_n$ can be written as

$$\begin{aligned}\mathrm{MMSE}\left(\mathbf{F}_{H,T+1}, \mathbf{Y}_{T+1}\right) &:= E\left(\left(\mathbf{Y}_{T+1} - \mathbf{F}_{H,T+1}\right)\left(\mathbf{Y}_{T+1} - \mathbf{F}_{H,T+1}\right)'\right) \\ &= \mathbf{H}^*\mathbf{V}\mathbf{H}^{*\prime} - \mathbf{H}^*\mathbf{W} - \mathbf{W}'\mathbf{H}^{*\prime} + \Sigma_{nn} \quad ,\end{aligned}$$

where $\mathbf{H}^* := [\mathbf{H}_1,...,\mathbf{H}_{n-1}] \sim k \times (n-1)k$.

To minimize $\mathrm{tr}\left(\mathrm{MMSE}\left(\mathbf{F}_{H,T+1}, \mathbf{Y}_{T+1}\right)\right)$ with repect to the restrictions (1) and (2), we consider:

$$L\left(\mathbf{H}^*, \lambda\right) := \mathrm{tr}\left(\mathbf{H}^*\mathbf{V}\mathbf{H}^{*\prime} - \mathbf{H}^*\mathbf{W} - \mathbf{W}'\mathbf{H}^{*\prime} + \Sigma_{nn}\right) - \lambda'\left(\mathbf{H}^*\gamma - \mu_n\right)$$

where $\lambda := (\lambda_1,...,\lambda_k)'$.

The necessary conditions for a minimum are:

I) $$\frac{\delta L(\mathbf{H}^*,\lambda)}{\delta \mathbf{H}^*} = 2\mathbf{H}^*\mathbf{V} - 2\mathbf{W}' - \lambda\gamma' \overset{!}{=} \mathbf{0}_{k\times(n-1)k}$$

II) $$\frac{\delta L(\mathbf{H}^*,\lambda)}{\delta \lambda'} = \mu_n{}' - (\mathbf{H}^*\gamma)' \overset{!}{=} \mathbf{0}_{1\times k}$$

From I) we get

$$\mathbf{H}^* = \mathbf{W}'\mathbf{V}^{-1} + \frac{1}{2}\lambda\gamma'\mathbf{V}^{-1}$$

and inserting in II) we obtain

$$\frac{1}{2}\lambda = \frac{1}{\gamma'\mathbf{V}^{-1}\gamma}\left(\mu_n - \mathbf{W}'\mathbf{V}^{-1}\gamma\right) \quad .$$

Back to I) results in

$$\mathbf{H}^*_{opt} = \mathbf{W}'\mathbf{V}^{-1} + \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\left(\mu_n - \mathbf{W}'\mathbf{V}^{-1}\gamma\right)\gamma'\mathbf{V}^{-1} \quad .$$

Using this weights for the combination and calculating the MMSE results in

$$\begin{aligned}\mathrm{MMSE}\left(\mathbf{F}_{H_{opt,T+1}}, \mathbf{Y}_{T+1}\right) = \Sigma_{nn} &- \mathbf{W}'\mathbf{V}^{-1}\mathbf{W} - \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\mathbf{W}'\mathbf{V}^{-1}\gamma\mu_n{}' - \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\mu_n\gamma'\mathbf{V}^{-1}\mathbf{W}\\ &+ \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\mu_n\mu_n{}' + \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\mathbf{W}'\mathbf{V}^{-1}\gamma\gamma'\mathbf{V}^{-1}\mathbf{W} \quad .\end{aligned}$$

Considering now an arbitrary $\mathbf{H}^*_{arb}$ satisfying the two restrictions, we can write $\mathbf{H}^*_{arb} := \tilde{\mathbf{W}}'\mathbf{V}^{-1} + \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\left(\mu_n - \tilde{\mathbf{W}}'\mathbf{V}^{-1}\gamma\right)\gamma'\mathbf{V}^{-1}$, where $\tilde{\mathbf{W}} \sim (n-1)k\times k$, and calculate

$$\begin{aligned}\mathrm{MMSE}\left(\mathbf{F}_{H_{arb,T+1}}, \mathbf{Y}_{T+1}\right) = \Sigma_{nn} &+ \tilde{\mathbf{W}}'\mathbf{V}^{-1}\tilde{\mathbf{W}} - \mathbf{W}'\mathbf{V}^{-1}\tilde{\mathbf{W}} - \tilde{\mathbf{W}}'\mathbf{V}^{-1}\mathbf{W} + \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\mu_n\mu_n{}'\\ &- \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\tilde{\mathbf{W}}'\mathbf{V}^{-1}\gamma\gamma'\mathbf{V}^{-1}\tilde{\mathbf{W}} - \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\mu_n\gamma'\mathbf{V}^{-1}\mathbf{W}\\ &+ \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\tilde{\mathbf{W}}'\mathbf{V}^{-1}\gamma\gamma'\mathbf{V}^{-1}\mathbf{W} - \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\mathbf{W}'\mathbf{V}^{-1}\gamma\mu_n{}'\\ &+ \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\mathbf{W}'\mathbf{V}^{-1}\gamma\gamma'\mathbf{V}^{-1}\tilde{\mathbf{W}} .\end{aligned}$$

Thus,

$$\mathrm{MMSE}\left(\mathbf{F}_{H_{arb,T+1}}, \mathbf{Y}_{T+1}\right) - \mathrm{MMSE}\left(\mathbf{F}_{H_{opt,T+1}}, \mathbf{Y}_{T+1}\right)$$

$$= \left(\tilde{\mathbf{W}}' - \mathbf{W}\right)\left(\mathbf{V}^{-1} - \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\mathbf{V}^{-1}\gamma\gamma'\mathbf{V}^{-1}\right)\left(\tilde{\mathbf{W}}' - \mathbf{W}'\right)'$$

and since $\left(\mathbf{V}^{-1} - \left(\gamma'\mathbf{V}^{-1}\gamma\right)^{-1}\mathbf{V}^{-1}\gamma\gamma'\mathbf{V}^{-1}\right)$ is n.n.d. (see e.g. Horn and Johnson, 1985, p. 47) the difference of the two MMSEs is n.n.d., as well.

In the multivariate case it is also possible to calculate bias based unbiased forecast combinations but because of the more complex bias structure there are several combination strategies. A simple procedure is to consider each component separately and then to derive the combination as we did in the univariate case, so that we get diagonal weight matrices. Similar to Theorem 2 we propose here:

Theorem 5: Let $\mathbf{F}_{i,T+1} := \left(F^{(i)}_{1,T+1},\dots,F^{(i)}_{k,T+1}\right)'$ be forecasts for $\mathbf{Y}_{T+1} := \left(Y_{1,T+1},\dots,Y_{k,T+1}\right)'$, $k \geq 2$, and $\mathbf{u}_{i,T+1} := \mathbf{Y}_{T+1} - \mathbf{F}_{i,T+1}$, where $E(\mathbf{u}_{i,T+1}) =: \mu_i$, $i = 1,\dots,n$. Further let $\mu_v = \sum_{\substack{i=1\\ i\neq v}}^{n} \mathbf{A}_i \mu_i$, where $v \in \{1,\dots,n\}$ is fixed but arbitrary, $\mathbf{A}_i \sim k \times k$, $i = 1,\dots,n,\ i \neq v$, and $\operatorname{rg}\left(\mathbf{I}_k - \sum_{\substack{i=1\\ i\neq v}}^{n} \mathbf{A}_i\right) = k$. Then

$$\mathbf{F}_{A,T+1} := \left(\mathbf{I}_k - \sum_{\substack{i=1\\ i\neq v}}^{n} \mathbf{A}_i\right)^{-1}\left(\mathbf{F}_{v,T+1} - \sum_{\substack{i=1\\ i\neq v}}^{n} \mathbf{A}_i \mathbf{F}_{i,T+1}\right)$$ is an unbiased forecast for $\mathbf{Y}_{T+1}$.

Proof: The mean of the error of the forecast combination is

$$E\left(\mathbf{Y}_{T+1} - \mathbf{F}_{A,T+1}\right) = E\left(\left(\mathbf{I}_k - \sum_{\substack{i=1\\ i\neq v}}^{n} \mathbf{A}_i\right)^{-1}\left(\left(\mathbf{I}_k - \sum_{\substack{i=1\\ i\neq v}}^{n} \mathbf{A}_i\right)\mathbf{Y}_{T+1} - \mathbf{F}_{v,T+1} + \sum_{\substack{i=1\\ i\neq v}}^{n} \mathbf{A}_i \mathbf{F}_{i,T+1}\right)\right)$$

$$= \left(\mathbf{I}_k - \sum_{\substack{i=1\\ i\neq v}}^{n} \mathbf{A}_i\right)^{-1} E\left(\mathbf{I}_k\left(\mathbf{Y}_{T+1} - \mathbf{F}_{v,T+1}\right) - \sum_{\substack{i=1\\ i\neq v}}^{n} \mathbf{A}_i\left(\mathbf{Y}_{T+1} - \mathbf{F}_{i,T+1}\right)\right)$$

$$= \left(\mathbf{I}_k - \sum_{\substack{i=1\\ i\neq v}}^{n} \mathbf{A}_i\right)^{-1}\left(\mathbf{I}_k \mu_v - \sum_{\substack{i=1\\ i\neq v}}^{n} \mathbf{A}_i \mu_i\right)$$

$$= \left(\mathbf{I}_k - \sum_{\substack{i=1\\ i\neq v}}^{n} \mathbf{A}_i\right)^{-1}\left(\mu_v - \mu_v\right) = \mathbf{0} \quad .$$

Remark: It is possible to use bias proportions in Theorem 5 for the definition of the matrices $\mathbf{A}_i$. If we assume that $\mu_{ij} \neq 0$, $i = 1,\dots,n,\ i \neq v,\ j = 1,\dots,k$, we get:

$$\mathbf{A}_i := \left(a_{rs}^{(i)}\right)_{r,s=1,\dots,k}, \text{ where } a_{rs}^{(i)} := \frac{1}{k(n-1)}\frac{\mu_{vr}}{\mu_{is}},\ i = 1,\dots,n, i \neq v,\ r,s = 1,\dots,k.$$

If we proceed in that way, we have to check if the assumption of regularity in Theorem 5 is satisfied.

Finally we present another general bias based combination method.

Theorem 6: Let $\mathbf{F}_{i,T+1} := \left(F_{1,T+1}^{(i)},\dots,F_{k,T+1}^{(i)}\right)'$ be forecasts for $\mathbf{Y}_{T+1} := \left(Y_{1,T+1},\dots,Y_{k,T+1}\right)'$, $k \geq 2$, and $\mathbf{u}_{i,T+1}$ the i-th forecast error vector, $\mu_i := (\mu_{i1},\dots,\mu_{ik})' := E(\mathbf{u}_{i,T+1})$, $i = 1,\dots,n$. Further let $\mu_{ij} \neq 0$, $i = 1,\dots,n,\ j = 1,\dots,k$. Then, $\mathbf{F}_{A,T+1} := \sum_{i=1}^{n} \mathbf{A}_i \mathbf{F}_{i,T+1}$ is an unbiased forecast for $\mathbf{Y}_{T+1}$, where

$$\mathbf{A}_i := \begin{pmatrix} a_{11}^{(i)} & \cdots & a_{1k}^{(i)} \\ \vdots & \ddots & \vdots \\ a_{k1}^{(i)} & \cdots & a_{kk}^{(i)} \end{pmatrix} \text{ and}$$

$$a_{rs}^{(i)} := \left(\frac{1}{n} Z_{rs} - \tilde{a}_{rs}^{(i)}\right) \Big/ M_r,\ Z_{rs} := \sum_{i=1}^{n} \tilde{a}_{rs}^{(i)} \quad , r,s = 1,..,k,\ r \neq s,$$

$$a_{rr}^{(i)} := -\frac{1}{\mu_{ir}} \sum_{\substack{j=1 \\ j\neq r}}^{k} a_{rj}^{(i)} \mu_{ij} \quad , i = 1,\dots,n,$$

$$M_r := \sum_{i=1}^{n} -\frac{1}{\mu_{ir}} \sum_{\substack{j=1 \\ j\neq r}}^{k} \left(\frac{1}{n} Z_{rj} - \tilde{a}_{rj}^{(i)}\right) \mu_{ij} \quad , r = 1,\dots,k$$

where the $\tilde{a}_{rs}^{(i)}$'s must be chosen so that $M_r \neq 0$.

Proof: For a fixed $h \in \{1,\dots,k\}$ we consider the h-th row of each of the n weight matrices, given by $\mathbf{a}_{h.}^{(i)} := \left(a_{h1}^{(i)},\dots,a_{hk}^{(i)}\right)'$, $i = 1,\dots,n$. Therefore,

$$\sum_{i=1}^{n} a_{hm}^{(i)} = \begin{cases} -\sum_{i=1}^{n} \frac{1}{\mu_{ih}} \sum_{\substack{p=1 \\ p\neq h}}^{k} a_{hp}^{(i)} \mu_{ip} & \text{if } m = h \\ \frac{1}{M_h} \left(\sum_{i=1}^{n} \frac{1}{n} Z_{hm} - \sum_{i=1}^{n} \tilde{a}_{hm}^{(i)} \right) & \text{if } m \in \{1,\ldots,k\},\ m \neq h \end{cases}$$

$$= \begin{cases} 1 & \text{if } m = h \\ 0 & \text{if } m \in \{1,\ldots,k\}, \quad m \neq h \end{cases} .$$

Thus we can write $Y_{h,T+1} = \sum_{i=1}^{n} \mathbf{a}_{h.}^{(i)} \mathbf{Y}_{T+1}$, and the mean of the combined forecast error in the h-th component is

$$E\left(\sum_{i=1}^{n} \mathbf{a}_{h.}^{(i)\prime} \mathbf{u}_{i,T+1} \right) = \sum_{i=1}^{n} a_{hh}^{(i)} \mu_{ih} + \sum_{i=1}^{n} \sum_{\substack{m=1 \\ m\neq h}}^{k} a_{hm}^{(i)} \mu_{im}$$

$$= -\sum_{i=1}^{n} \left(\frac{1}{\mu_{ih}} \sum_{\substack{m=1 \\ m\neq h}}^{k} a_{hm}^{(i)} \mu_{im} \right) \mu_{ih} + \sum_{i=1}^{n} \sum_{\substack{m=1 \\ m\neq h}}^{k} a_{hm}^{(i)} \mu_{im} = 0 \ .$$

If we look at Theorem 6 again, we have to notice that the $\tilde{a}_{rs}^{(i)}$'s are not specified there. The practitioner could choose them by his subjective view of the given problem. Obviously such a general method could also be defined for the univariate case, but because of the subjective choice of the $\tilde{a}_{rs}^{(i)}$'s , this is excluded from the simulation study and therefore not presented in Section 2.1.

3. References

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

[2] **Diebold, F.X., Pauly, P.** (1987): "Structural change and the combination of forecasts", *Journal of Forecasting 6*, 21-40.

[3] **Deutsch, M., Granger C.W.J., Teräsvirta, T.** (1994): "The combination of forecasts using changing weights", *International Journal of Forecasting 10*, 47-57.

[4] **Ehrbeck, T.**, **Waldmann, R.** (1996): "Why are professional forecasters biased? Agency versus behavioral explanations", *The Quarterly Journal of Economics*, 21-40, 1996.

[5] **Granger, C.W.J., Ramanathan, R.** (1984): "Improved methods of combining forecasts", *Journal of Forecasting 3*, 197-204.

[6] **Horn, R.A., Johnson, R.J.** (1985): "Matrix analysis", *Cambridge University Press*, 1985.

[7] **Klapper, M.** (1998): "Combining German macro economic forecasts using rank-based techniques", *Technical Report 19/1998. University of Dortmund.*

[8] **Odell, P.L., Dorsett, D., Young, D., Igwe, J.** (1989): "Estimator models for combining vector estimators", *Mathematical Computational Modelling 12*, 1627-1642.

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