Scalar Mean Square Error Optimal Linear Combination of Multivariate Forecasts

Sven-Oliver Troschke
Department of Statistics, University of Dortmund,
44221 Dortmund, Germany
troschke@statistik.uni-dortmund.de

Abstract: When a forecaster predicts the future value of a certain random variable it is very likely that he will not only forecast that certain variable but he will also forecast other variables from the same field. In the literature on the combination of several individual forecasts univariate approaches have been used almost exclusively. They deal with each forecasted variable at a time. In doing so all the information stemming from the interaction of the variables is neglected. The aim of this report is to show how a set of such *multivariate forecasts* can be combined efficiently. We will focus on various linear combinations and determine how the combination weights should be chosen optimally with respect to the scalar mean square prediction error (SMSPE) criterion. For this purpose we will assume that the first and second order moments of the joint distribution of target variable and individual forecasts are given. As a by-product linear adjustments of single forecasts are obtained. An example illustrating the potential inherent in the multivariate approaches compared to the classical univariate methods is presented. The performance of these methods has to be reassessed if the moments of the joint distribution are unknown and have to be estimated. Further investigations have to be carried out.

Keywords: Combination of forecasts, multivariate forecasts, linear combination.

AMS 2000 Subject Classification: 62M20

1 Introduction

Let us be given k forecasts $\mathbf{f}_1, \ldots, \mathbf{f}_k$ for an l-dimensional random vector $\mathbf{y} = (y_1, \ldots, y_l)^{\mathrm{T}}$. The forecasts are stacked to form a random vector $\mathbf{f} \sim (kl, 1)$, i.e.

$\mathbf{f} = (\mathbf{f}_1^{\mathrm{T}}, \dots, \mathbf{f}_k^{\mathrm{T}})^{\mathrm{T}}$. Our objective is to obtain a combined forecast vector $\mathbf{f}_{\text{comb}}$ from the single forecasts $\mathbf{f}_i$ aiming at optimality within certain given classes of linear combinations.

In the literature on the combination of forecasts univariate approaches have been used almost exclusively, dealing with each forecasted variable y_j at a time. Only few work can be found on multivariate forecasts: FUHRER and HALTMAIER (1988) and WENZEL (1999b, 2001) state the MMSPE-optimal choice of the combination weights in a special case. KLAPPER (1999, 2000) develops rank-based procedures for the combination of multivariate forecasts, whereas WENZEL (1998, 1999a, 2001) determines optimal combination weights on the basis of a multivariate Pitman-closeness criterion. Finally, WENZEL (2000, 2001) investigates the effect of shrinking combined forecasts.

When applying univariate combinations, the information stemming from the interaction of the variables is neglected. The aim of this paper is to provide an extensive analysis on how a set of multivariate forecasts can be combined effectively. We will focus on various linear combinations and determine how the combination parameters should be chosen optimally with respect to the scalar mean square prediction error criterion.

A multivariate linear combination of the forecasts $\mathbf{f}_1, \dots, \mathbf{f}_k$ is given by

$$\mathbf{f}_{\text{comb}} = \mathbf{B}_1\mathbf{f}_1 + \ldots + \mathbf{B}_k\mathbf{f}_k + \mathbf{c} = \sum_{i=1}^{k} \mathbf{B}_i\mathbf{f}_i + \mathbf{c} = \mathbf{B}\mathbf{f} + \mathbf{c}\,, \tag{1.1}$$

with $\mathbf{B} = (\mathbf{B}_1 \vdots \ \ldots \vdots \ \mathbf{B}_k) \in \mathbb{R}^{l\times kl}$ and $\mathbf{c} \in \mathbb{R}^l$.

The number of parameters involved in such a linear combination is quite large, namely $kl^2 + l$. Consequently, it may be worthwhile considering variants which employ a smaller number of parameters but still capture the spirit of a multivariate combination. This is especially true if the combination parameters have to be estimated from empirical data.

But there are other reasons as well for which it may be appropriate to place certain restrictions on the combination parameters $\mathbf{B}$ and $\mathbf{c}$. For example, if all the individual forecasts $\mathbf{f}_i$ are unbiased for $\mathbf{y}$, i.e. $\mathrm{E}(\mathbf{f}_i - \mathbf{y}) = \mathbf{0}$, $i = 1, \dots, k$, the combined forecast will be unbiased as well if we restrict $\mathbf{c} = \mathbf{0}$ and $\sum_{i=1}^k \mathbf{B}_i = \mathbf{I}_l$.

Concerning optimality it is important to say by what measure the quality of a forecast $\tilde{\mathbf{f}}$ is to be judged. The most obvious measure is the matrix mean square prediction error

$$\begin{aligned} \mathrm{MMSPE}(\tilde{\mathbf{f}}, \mathbf{y}) &= \mathrm{E}[(\mathbf{y} - \tilde{\mathbf{f}})(\mathbf{y} - \tilde{\mathbf{f}})^{\mathrm{T}}] \\ &= \mathrm{Cov}(\mathbf{y} - \tilde{\mathbf{f}}) + [\mathrm{E}(\mathbf{y} - \tilde{\mathbf{f}})][\mathrm{E}(\mathbf{y} - \tilde{\mathbf{f}})]^{\mathrm{T}}\,. \end{aligned} \tag{1.2}$$

An alternative is offered by its scalar counterpart, the scalar mean square prediction error

$$\begin{aligned}\mathrm{SMSPE}(\tilde{\mathbf{f}},\mathbf{y}) &= \mathrm{E}[(\mathbf{y}-\tilde{\mathbf{f}})^{\mathrm{T}}(\mathbf{y}-\tilde{\mathbf{f}})] \\ &= \mathrm{tr}(\mathrm{Cov}(\mathbf{y}-\tilde{\mathbf{f}})) + [\mathrm{E}(\mathbf{y}-\tilde{\mathbf{f}})]^{\mathrm{T}}[\mathrm{E}(\mathbf{y}-\tilde{\mathbf{f}})] \\ &= \mathrm{tr}(\mathrm{MMSPE}(\tilde{\mathbf{f}},\mathbf{y}))\ . \end{aligned} \tag{1.3}$$

The SMSPE is a scalar valued function. Hence, comparison of several forecasts is more easily accomplished than by using the MMSPE which is matrix valued. Here comparisons would have to be carried out in the LÖWNER ordering (LÖWNER, 1934). Consequently, SMSPE-optimality will be our target criterion. For the linear combinations involving a full parameter matrix $\mathbf{B}$, however, also optimality with respect to the matrix valued MMSPE-criterion is granted, as we will demonstrate in the respective sections. This corresponds to a result by ODELL, DORSETT, YOUNG and IGWE (1989) who investigate the linear combination of vector estimators.
So, we will identify the SMSPE-optimal choice for the respective combination parameters. As we will see later on the optimal choice requires different levels of knowledge about the moments of the joint distribution of $\mathbf{y}$ and $\mathbf{f}$ depending on the chosen variant of the linear combination. In each case, however, moments up to order two are involved. We will now introduce our notations:
Generalizing the approach from HARVILLE (1985) to the case of multivariate forecasts we will assume the following setting: The expectations of $\mathbf{y}$ and $\mathbf{f}$ are given by $\mathrm{E}(\mathbf{y}) = \boldsymbol{\mu}_0$ and $\mathrm{E}(\mathbf{f}) = \mathrm{E}((\mathbf{f}_1^{\mathrm{T}},\dots,\mathbf{f}_k^{\mathrm{T}})^{\mathrm{T}}) = (\boldsymbol{\mu}_1^{\mathrm{T}},\dots,\boldsymbol{\mu}_k^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{\mu}_{\mathbf{f}}$ giving rise to the model:

$$\begin{pmatrix}\mathbf{y}\\ \mathbf{f}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\mu}_0\\ \boldsymbol{\mu}_{\mathbf{f}}\end{pmatrix} + \begin{pmatrix}\boldsymbol{\varepsilon}_0\\ \boldsymbol{\varepsilon}_{\mathbf{f}}\end{pmatrix} =: \boldsymbol{\mu} + \boldsymbol{\varepsilon}\ , \tag{1.4}$$

where $\boldsymbol{\varepsilon}_{\mathbf{f}} := (\boldsymbol{\varepsilon}_1^{\mathrm{T}},\dots,\boldsymbol{\varepsilon}_k^{\mathrm{T}})^{\mathrm{T}}$. Consequently, $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and the higher order moments of $\boldsymbol{\varepsilon}$ are the centered moments of $(y,\mathbf{f}^{\mathrm{T}})^{\mathrm{T}}$.
The elements of the l-dimensional target vector variable $\mathbf{y}$ are denoted by $\mathbf{y} = (y_1,\dots,y_l)^{\mathrm{T}}$, whereas for $i = 1,\dots,k$ the elements of the forecast vector $\mathbf{f}_i$ are denoted by $\mathbf{f}_i = (f_{i,1},\dots,f_{i,l})^{\mathrm{T}}$. The elements of the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\varepsilon}$ are named accordingly.
The second order moments are given by

$$\boldsymbol{\Sigma} := \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = \mathrm{E}\left[\begin{pmatrix}\boldsymbol{\varepsilon}_0\\ \boldsymbol{\varepsilon}_{\mathbf{f}}\end{pmatrix}\begin{pmatrix}\boldsymbol{\varepsilon}_0\\ \boldsymbol{\varepsilon}_{\mathbf{f}}\end{pmatrix}^{\mathrm{T}}\right] =: \begin{pmatrix}\boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{0\mathbf{f}}\\ \boldsymbol{\Sigma}_{\mathbf{f}0} & \boldsymbol{\Sigma}_{\mathbf{ff}}\end{pmatrix} \tag{1.5}$$

and

$$\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = \mathrm{E}\left[\left(\begin{pmatrix} \mathbf{y} \\ \mathbf{f} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu}_0 \\ \boldsymbol{\mu}_{\mathbf{f}} \end{pmatrix}\right)\left(\begin{pmatrix} \mathbf{y} \\ \mathbf{f} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu}_0 \\ \boldsymbol{\mu}_{\mathbf{f}} \end{pmatrix}\right)^{\mathrm{T}}\right] = \mathrm{Cov}\left[\begin{pmatrix} \mathbf{y} \\ \mathbf{f} \end{pmatrix}\right] . \tag{1.6}$$

Naturally, $\boldsymbol{\Sigma}_{00}$, $\boldsymbol{\Sigma}_{\mathbf{ff}}$ and $\boldsymbol{\Sigma}$ are symmetric nonnegative definite matrices.
The lower left $(kl \times l)$-submatrix $\boldsymbol{\Sigma}_{\mathbf{f}0}$ and the lower right $(kl \times kl)$-submatrix $\boldsymbol{\Sigma}_{\mathbf{ff}}$ of $\boldsymbol{\Sigma}$ are block matrices consisting of $(l \times l)$-dimensional blocks:

$$\boldsymbol{\Sigma}_{\mathbf{f}0} = \begin{pmatrix} \boldsymbol{\Sigma}_{10} \\ \boldsymbol{\Sigma}_{20} \\ \vdots \\ \boldsymbol{\Sigma}_{k0} \end{pmatrix} \quad \text{and } \boldsymbol{\Sigma}_{\mathbf{ff}} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \dots & \boldsymbol{\Sigma}_{1k} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \dots & \boldsymbol{\Sigma}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{k1} & \boldsymbol{\Sigma}_{k2} & \dots & \boldsymbol{\Sigma}_{kk} \end{pmatrix} . \tag{1.7}$$

We will assume invertibility of the centered second order moment matrix of $\mathbf{f}$ throughout, i.e. we assume invertibility of $\boldsymbol{\Sigma}_{\mathbf{ff}} = \mathrm{Cov}(\mathbf{f})$, and hence also invertibility of the non-centered second order moment matrix $\boldsymbol{\Sigma}_{\mathbf{ff}} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}} = \mathrm{E}(\mathbf{f}\mathbf{f}^{\mathrm{T}})$ is granted. Note that vectors and matrices are represented by bold face letters.
For the determination of the optimal combination parameters we will assume that the first and second order moments of the joint distribution of $\mathbf{y}$ and $\mathbf{f}$ exist. If this is not the case, e.g. if a component y_j of the target vector variable $\mathbf{y}$ is trended, then appropriate transformations of $\mathbf{y}$ and $\mathbf{f}$ should be undertaken, e.g. differencing of the time series of observations on y_j or consideration of relative changes. Since $\mathbf{f}_1, \dots, \mathbf{f}_k$ are forecasts of $\mathbf{y}$ the same transformation should work for both, target variable and forecasts.
Furthermore, we will assume that we *know* the first and second order moments of the joint distribution of $\mathbf{y}$ and $\mathbf{f}$, but not the distribution itself. (This describes state 2 of knowledge in the classification scheme by HARVILLE (1985). State 1 means complete knowledge about the distribution.)
We will see that the variants of the SMSPE-optimal combined forecasts depend on different portions of these first and second order moments. In practical applications such moments will hardly ever be known. (Thus our knowledge falls even behind state 4 of knowledge in HARVILLE's scheme, where some assumptions on the first order moments are made.) Consequently, we will have to estimate the necessary moments from a sample of observations on the variables of interest. Then we plug these estimators into the formulae for the optimal combinations. We may apply the ordinary sample moments as estimators, but of course one might think of using alternatives for this step, e.g. robust estimators.

Section 2 deals with the basic multivariate linear approaches, the univariate linear approaches can be obtained as the special case $l = 1$. An alternative view on two of the linear approaches via consideration of forecast errors is presented in Section 3. Section 4 investigates variants which require less knowledge about the covariance structure of the joint distribution of $\mathbf{y}$ and $\mathbf{f}$. Section 5 considers the special case of $k = 1$ forecast, which results in adjustment of an individual forecast. Section 6 presents an analysis where for a set of exemplary first and second order moments of the joint distribution of $\mathbf{y}$ and $\mathbf{f}$ the potential of the various multivariate adjustments and combinations of forecasts is explored and compared to the univariate treatment of each variable involved. The question in how far the various methods are sensitive to the chosen coordinate system is discussed in Section 7. Section 8 concludes the report. Appendix A lists some results mostly from the theory of matrix differential calculus which will be useful in the subsequent sections. Appendix B proves Lemma 3.1.

We are now going to investigate the linear approach to the combination of multivariate forecasts.

2 The multivariate linear approach

In the univariate case the linear combination approach is used predominantly in the literature. A good overview on the many investigations carried out in this direction is provided e.g. by Clemen (1989) or Thiele (1993). Hence, it is nearby to concentrate on *linear* combinations of multivariate forecasts in the first place. The results for the linear combination of univariate forecasts follow as the special case $l = 1$ from the subsequent derivations (compare also Troschke and Trenkler, 2000).

Linearly combined forecasts are of the form

$$\mathbf{B}\mathbf{f} + \mathbf{c} = \mathbf{B}_1\mathbf{f}_1 + \ldots + \mathbf{B}_k\mathbf{f}_k + \mathbf{c} = \sum_{i=1}^{k} \mathbf{B}_i\mathbf{f}_i + \mathbf{c} \; , \tag{2.1}$$

where it may be appropriate to place certain restrictions on the combination parameters $\mathbf{B} = (\mathbf{B}_1|\mathbf{B}_2|\ldots|\mathbf{B}_k) \in \mathbb{R}^{l\times kl}$ and $\mathbf{c} \in \mathbb{R}^l$.

We will consider two possible restrictions on the combination parameters. First we may want to neglect the constant term $\mathbf{c}$ in the linear combination, i.e. we may use the restriction $\mathbf{c} = \mathbf{0}$. On the other hand we may restrict the parameter matrices such that they sum up to unity, i.e. $\sum_{i=1}^{k} \mathbf{B}_i = \mathbf{I}_l$, where $\mathbf{I}_l$ is the $l \times l$ identity matrix. As mentioned in the introduction using both restrictions results in an unbiased combination from unbiased individual forecasts.

Since each of the two restrictions may or may not be utilized we arrive at four different linear combinations, which we are now going to examine. For each of the variants we will state how the combination parameters should be chosen in order to minimize the scalar mean square prediction error of such a combined forecast and we will provide the respective minimal values.

The first considered variant of the linear approach is

$$f_{\mathbf{B},\mathbf{c}} = \mathbf{B}\mathbf{f} + \mathbf{c} \; . \tag{2.2}$$

No restrictions are imposed on the combination parameters $\mathbf{B}$ and $\mathbf{c}$ in this set-up. The expectation of $\mathbf{f}_{\mathbf{B},\mathbf{c}}$ is given as

$$\mathrm{E}(\mathbf{f}_{\mathbf{B},\mathbf{c}}) = \mathrm{E}(\mathbf{B}\mathbf{f} + \mathbf{c}) = \mathbf{B}\,\mathrm{E}(\mathbf{f}) + \mathbf{c} = \mathbf{B}\boldsymbol{\mu}_{\mathbf{f}} + \mathbf{c} \; . \tag{2.3}$$

We now want to determine how the combination parameters $\mathbf{B}$ and $\mathbf{c}$ should be chosen in order to minimize the scalar mean square prediction error of such a combined forecast. To achieve this goal we will perform the following three steps: In the first step we will explicitly calculate the SMSPE-function. With the help of matrix differential calculus we will differentiate this function with respect to the combination parameters $\mathbf{B}$ and $\mathbf{c}$ in the second step. In the final step we will simultaneously equate these derivatives to zero and solve the resulting linear equation system. (Compare also Definition A.3 and Lemma A.4.) The unique solution $(\mathbf{B}_{\mathrm{opt}}, \mathbf{c}_{\mathrm{opt}})$ of the equation system produces the desired minimum of the SMSPE-function.

Step 1: Explicit calculation of the SMSPE**-function.** Since it will be useful later on we first calculate the matrix mean square prediction error function and then exploit the fact that $\mathrm{SMSPE}(\mathbf{f}_{\mathbf{B},\mathbf{c}}, \mathbf{y}) = \mathrm{tr}(\mathrm{MMSPE}(\mathbf{f}_{\mathbf{B},\mathbf{c}}, \mathbf{y}))$:

$$\begin{aligned} \mathrm{MMSPE}(\mathbf{f}_{\mathbf{B},\mathbf{c}}, \mathbf{y}) &= \mathrm{E}[(\mathbf{y} - \mathbf{f}_{\mathbf{B},\mathbf{c}})(\mathbf{y} - \mathbf{f}_{\mathbf{B},\mathbf{c}})^{\mathrm{T}}] \\ &= \mathrm{Cov}(\mathbf{y} - \mathbf{f}_{\mathbf{B},\mathbf{c}}) + [\mathrm{E}(\mathbf{y} - \mathbf{f}_{\mathbf{B},\mathbf{c}})][\mathrm{E}(\mathbf{y} - \mathbf{f}_{\mathbf{B},\mathbf{c}})]^{\mathrm{T}} \\ &= \boldsymbol{\Sigma}_{00} - \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{f}0} - \boldsymbol{\Sigma}_{0\mathbf{f}}\mathbf{B}^{\mathrm{T}} + \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{ff}}\mathbf{B}^{\mathrm{T}} \\ &\quad + (\boldsymbol{\mu}_0 - \mathbf{B}\boldsymbol{\mu}_{\mathbf{f}} - \mathbf{c})(\boldsymbol{\mu}_0 - \mathbf{B}\boldsymbol{\mu}_{\mathbf{f}} - \mathbf{c})^{\mathrm{T}} \; . \end{aligned} \tag{2.4}$$

Taking traces and regrouping the terms with respect to the occurring unknowns we arrive at the following expression for the scalar mean square prediction error of $\mathbf{f}_{\mathbf{B},\mathbf{c}}$:

$$\begin{aligned}
&\mathrm{SMSPE}(\mathbf{f}_{\mathbf{B},\mathbf{c}},\mathbf{y}) = \mathrm{tr}(\mathrm{MMSPE}(\mathbf{f}_{\mathbf{B},\mathbf{c}},\mathbf{y})) \\
&= \quad \mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{B}(\boldsymbol{\Sigma}_{\mathbf{ff}} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}})) \\
&\quad - 2\,\mathrm{tr}(\mathbf{B}(\boldsymbol{\Sigma}_{\mathbf{f}0} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{0}^{\mathrm{T}})) \\
&\quad + 2\mathbf{c}^{\mathrm{T}}\mathbf{B}\boldsymbol{\mu}_{\mathbf{f}} \\
&\quad + \mathbf{c}^{\mathrm{T}}\mathbf{c} \\
&\quad - 2\boldsymbol{\mu}_{0}^{\mathrm{T}}\mathbf{c} \\
&\quad + \mathrm{tr}(\boldsymbol{\Sigma}_{00}) + \boldsymbol{\mu}_{0}^{\mathrm{T}}\boldsymbol{\mu}_{0} \ . \qquad (2.5)
\end{aligned}$$

Step 2: Differentiation. Applying Lemma (A.5) we get

$$\frac{\partial \mathrm{SMSPE}(\mathbf{f}_{\mathbf{B},\mathbf{c}},\mathbf{y})}{\partial \mathbf{c}} = 2\left[\mathbf{c} - \boldsymbol{\mu}_0 + \mathbf{B}\boldsymbol{\mu}_{\mathbf{f}}\right] \ . \qquad (2.6)$$

Lemma A.6 leads us to

$$\frac{\partial \mathrm{SMSPE}(\mathbf{f}_{\mathbf{B},\mathbf{c}},\mathbf{y})}{\partial \mathbf{B}} \quad = \quad 2\left[\mathbf{B}(\boldsymbol{\Sigma}_{\mathbf{ff}} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}) - (\boldsymbol{\Sigma}_{0\mathbf{f}} + \boldsymbol{\mu}_0\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}) + \mathbf{c}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}\right] \ . \qquad (2.7)$$

Step 3: Equating to zero. Setting Equations (2.6) and (2.7) simultaneously to zero and solving the resulting linear equation system for the unknown parameters we obtain the optimal choices for $\mathbf{B}$ and $\mathbf{c}$.
From Equation (2.6) we get

$$\mathbf{c}_{\mathrm{opt}} = \boldsymbol{\mu}_0 - \mathbf{B}_{\mathrm{opt}}\boldsymbol{\mu}_{\mathbf{f}} \ . \qquad (2.8)$$

Using (2.8) from (2.7) we obtain

$$\mathbf{B}_{\mathrm{opt}} = \boldsymbol{\Sigma}_{0\mathbf{f}}\boldsymbol{\Sigma}_{\mathbf{ff}}^{-1} \ . \qquad (2.9)$$

Inserting this result into Equation (2.8) we finally arrive at

$$\mathbf{c}_{\mathrm{opt}} \quad = \quad \boldsymbol{\mu}_0 - \boldsymbol{\Sigma}_{0\mathbf{f}}\boldsymbol{\Sigma}_{\mathbf{ff}}^{-1}\boldsymbol{\mu}_{\mathbf{f}} \ . \qquad (2.10)$$

From Equation (2.3) it is obvious that the combined forecast $\mathbf{f}_{\mathbf{B}_{\mathrm{opt}},\mathbf{c}_{\mathrm{opt}}}$ is unbiased even if the single forecasts are biased.
Provided that the solution $(\mathbf{B}_{\mathrm{opt}}, \mathbf{c}_{\mathrm{opt}})$ is unique, it can be seen that this solution describes a minimum of the SMSPE-function within the considered class of combined

forecasts:

$$\begin{aligned} \text{SMSPE}(\mathbf{f}_{\mathbf{B},\mathbf{c}}, \mathbf{y}) &= \text{E}[(\mathbf{y} - \mathbf{f}_{\mathbf{B},\mathbf{c}})^{\text{T}}(\mathbf{y} - \mathbf{f}_{\mathbf{B},\mathbf{c}})] \\ &= \text{E}[(\mathbf{y} - \mathbf{Bf} - \mathbf{c})^{\text{T}}(\mathbf{y} - \mathbf{Bf} - \mathbf{c})] \\ &= \text{E}\left[\sum_{j=1}^{l} ((\mathbf{y} - \mathbf{Bf} - \mathbf{c})_j)^2\right] \\ &= \text{E}\left[\sum_{j=1}^{l} (y_j - \mathbf{B}_{j\cdot}\mathbf{f} - c_j)^2\right] \end{aligned} \tag{2.11}$$

is a quadratic function in the unknown parameters bounded below by the value 0. Here $(\mathbf{y} - \mathbf{Bf} - \mathbf{c})_j$ denotes the j-th component of the vector $\mathbf{y} - \mathbf{Bf} - \mathbf{c}$ and $B_{j\cdot} = (B_{j1}, \ldots, B_{j(kl)})$ denotes the j-th row of the parameter matrix $\mathbf{B} \in \mathbb{R}^{l \times kl}$.
Inserting $(\mathbf{B}_{\text{opt}}, \mathbf{c}_{\text{opt}})$ into Equation (2.5) we may derive that the optimal SMSPE-value is given by

$$\text{SMSPE}(\mathbf{f}_{\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}}}, \mathbf{y}) = \text{tr}(\boldsymbol{\Sigma}_{00}) - \text{tr}(\boldsymbol{\Sigma}_{0\mathbf{f}}\boldsymbol{\Sigma}_{\mathbf{ff}}^{-1}\boldsymbol{\Sigma}_{\mathbf{f}0}) \;. \tag{2.12}$$

As indicated in the introduction the above choice of combination parameters is not only optimal with respect to the SMSPE- but also with respect to the MMSPE-criterion: By inserting $\mathbf{B}_{\text{opt}}$ and $\mathbf{c}_{\text{opt}}$ into Equation (2.4) we get the corresponding value

$$\text{MMSPE}(\mathbf{f}_{\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}}}, \mathbf{y}) = \boldsymbol{\Sigma}_{00} - \boldsymbol{\Sigma}_{0\mathbf{f}}\boldsymbol{\Sigma}_{\mathbf{ff}}^{-1}\boldsymbol{\Sigma}_{\mathbf{f}0} \;. \tag{2.13}$$

Using Equations (2.4) and (2.13) we can see that the difference

$$\begin{aligned} &\text{MMSPE}(\mathbf{f}_{\mathbf{B},\mathbf{c}}, \mathbf{y}) - \text{MMSPE}(\mathbf{f}_{\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}}}, \mathbf{y}) \\ &= (\boldsymbol{\mu}_0 - \mathbf{B}\boldsymbol{\mu}_{\mathbf{f}} - \mathbf{c})(\boldsymbol{\mu}_0 - \mathbf{B}\boldsymbol{\mu}_{\mathbf{f}} - \mathbf{c})^{\text{T}} \\ &\quad + \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{ff}}\mathbf{B}^{\text{T}} - \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{f}0} - \boldsymbol{\Sigma}_{0\mathbf{f}}\mathbf{B}^{\text{T}} + \boldsymbol{\Sigma}_{0\mathbf{f}}\boldsymbol{\Sigma}_{\mathbf{ff}}^{-1}\boldsymbol{\Sigma}_{\mathbf{f}0} \\ &= (\boldsymbol{\mu}_0 - \mathbf{B}\boldsymbol{\mu}_{\mathbf{f}} - \mathbf{c})(\boldsymbol{\mu}_0 - \mathbf{B}\boldsymbol{\mu}_{\mathbf{f}} - \mathbf{c})^{\text{T}} \\ &\quad + (\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{ff}} - \boldsymbol{\Sigma}_{0\mathbf{f}})\boldsymbol{\Sigma}_{\mathbf{ff}}^{-1}(\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{ff}} - \boldsymbol{\Sigma}_{0\mathbf{f}})^{\text{T}} \end{aligned} \tag{2.14}$$

is always nonnegative definite, such that MMSPE-optimality is shown.
For a simple example see Section 6 where the potential of this combination is investigated for exemplary choices of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in the case of $k = 2$ forecasts of dimension $l = 2$.

If we drop the constant term $\mathbf{c}$ and still place no restrictions on $\mathbf{B}$, we obtain

$$f_{\mathbf{B}} = \mathbf{Bf} \;, \tag{2.15}$$

with expectation

$$\mathrm{E}(f_{\mathbf{B}}) = \mathbf{B}\boldsymbol{\mu}_{\mathbf{f}} \ . \tag{2.16}$$

Setting $\mathbf{c} = \mathbf{0}$ we derive the SMSPE of $f_{\mathbf{B}}$ from Equation (2.5):

$$\begin{aligned} &\mathrm{SMSPE}(\mathbf{f}_{\mathbf{B}}, \mathbf{y}) = \\ &\quad \mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{B}(\boldsymbol{\Sigma}_{\mathbf{ff}} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}})) \\ &\quad - 2\,\mathrm{tr}(\mathbf{B}(\boldsymbol{\Sigma}_{\mathbf{f}0} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{0}^{\mathrm{T}})) \\ &\quad + \mathrm{tr}(\boldsymbol{\Sigma}_{00}) + \boldsymbol{\mu}_{0}^{\mathrm{T}}\boldsymbol{\mu}_{0} \ . \end{aligned} \tag{2.17}$$

Differentiation with respect to $\mathbf{B}$ and setting the derivative equal to zero delivers the optimal choice for $\mathbf{B}$ within this approach:

$$\mathbf{B}_{\mathrm{opt}} = (\boldsymbol{\Sigma}_{0\mathbf{f}} + \boldsymbol{\mu}_{0}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}})\,(\boldsymbol{\Sigma}_{\mathbf{ff}} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}})^{-1} \ . \tag{2.18}$$

It is obvious that $\mathbf{f}_{\mathbf{B}_{\mathrm{opt}}}$ is not necessarily unbiased even if the individual forecasts are unbiased. The corresponding optimal SMSPE-value is given by

$$\begin{aligned} \mathrm{SMSPE}(\mathbf{f}_{\mathbf{B}_{\mathrm{opt}}}, \mathbf{y}) \ &= \ \mathrm{tr}(\boldsymbol{\Sigma}_{00}) + \boldsymbol{\mu}_{0}^{\mathrm{T}}\boldsymbol{\mu}_{0} \\ &- \ \mathrm{tr}\left[(\boldsymbol{\Sigma}_{0\mathbf{f}} + \boldsymbol{\mu}_{0}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}})(\boldsymbol{\Sigma}_{\mathbf{ff}} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}})^{-1}(\boldsymbol{\Sigma}_{\mathbf{f}0} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{0}^{\mathrm{T}})\right] \ . \end{aligned} \tag{2.19}$$

With the help of Lemma A.1 this may be rewritten as

$$\begin{aligned} \mathrm{SMSPE}(\mathbf{f}_{\mathbf{B}_{\mathrm{opt}}}, \mathbf{y}) \ &= \ \mathrm{tr}(\boldsymbol{\Sigma}_{00}) - \mathrm{tr}(\boldsymbol{\Sigma}_{0\mathbf{f}}\boldsymbol{\Sigma}_{\mathbf{ff}}^{-1}\boldsymbol{\Sigma}_{\mathbf{f}0}) \\ &+ \ \frac{\left(\boldsymbol{\mu}_{0} - \boldsymbol{\Sigma}_{0\mathbf{f}}\boldsymbol{\Sigma}_{\mathbf{ff}}^{-1}\boldsymbol{\mu}_{\mathbf{f}}\right)^{\mathrm{T}}\left(\boldsymbol{\mu}_{0} - \boldsymbol{\Sigma}_{0\mathbf{f}}\boldsymbol{\Sigma}_{\mathbf{ff}}^{-1}\boldsymbol{\mu}_{\mathbf{f}}\right)}{1 + \boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{ff}}^{-1}\boldsymbol{\mu}_{\mathbf{f}}} \end{aligned} \tag{2.20}$$

such that in view of Equation (2.12) the loss caused by dropping the constant term becomes obvious.
Similarly to the previous combination we can show that the difference

$$\begin{aligned} &\mathrm{MMSPE}(\mathbf{f}_{\mathbf{B}}, \mathbf{y}) - \mathrm{MMSPE}(\mathbf{f}_{\mathbf{B}_{\mathrm{opt}}}, \mathbf{y}) \\ &\ = \ [\mathbf{B}(\boldsymbol{\Sigma}_{\mathbf{ff}} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}) - (\boldsymbol{\Sigma}_{0\mathbf{f}} + \boldsymbol{\mu}_{0}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}})] \cdot \\ &\qquad \cdot [\boldsymbol{\Sigma}_{\mathbf{ff}} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}]^{-1}[\mathbf{B}(\boldsymbol{\Sigma}_{\mathbf{ff}} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}) - (\boldsymbol{\Sigma}_{0\mathbf{f}} + \boldsymbol{\mu}_{0}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}})]^{\mathrm{T}} \end{aligned} \tag{2.21}$$

is always nonnegative definite, such that $\mathbf{f}_{\mathbf{B}_{\mathrm{opt}}}$ is the MMSPE-optimal combination as well.

The remaining two combinations $\mathbf{f}_{\mathbf{B},\mathbf{c},\mathrm{rest}} = \mathbf{B}\mathbf{f} + \mathbf{c}$ and $\mathbf{f}_{\mathbf{B},\mathrm{rest}} = \mathbf{B}\mathbf{f}$ utilize the restriction $\sum_{i=1}^{k} \mathbf{B}_i = \mathbf{I}_l$ of the weight matrices summing up to the identity matrix.

Like in the univariate case (compare Troschke and Trenkler, 2000, Section 2) it is possible to incorporate this restriction into the goal function and then calculate the optimal combination parameters on that basis. The process, however, gets rather involved. To avoid this the next section will present an alternative view on forecast combination based on the forecast errors. More importantly, this alternative view demonstrates the close relationship with the covariance adjustment technique introduced by Rao (1966, 1967). Note that the restriction $\sum_{i=1}^{k} \mathbf{B}_i = \mathbf{I}_l$ is essential for this alternative view.

3 Alternative representations using error terms

Let us first consider the combined forecast

$$\mathbf{f}_{\mathbf{B},\text{rest}} = \mathbf{B}\mathbf{f} = \sum_{i=1}^{k} \mathbf{B}_i \mathbf{f}_i \quad \text{with} \quad \sum_{i=1}^{k} \mathbf{B}_i = \mathbf{I}_l \ . \tag{3.1}$$

This combination is designed for the case where each single forecast $\mathbf{f}_i$ is unbiased and hence $\boldsymbol{\mu}_i = \mathrm{E}(\mathbf{f}_i) = \mathrm{E}(\mathbf{y}) = \boldsymbol{\mu}_0$ for $i = 1, \ldots, k$ is assumed in the calculation of the optimal combination weights for $\mathbf{f}_{\mathbf{B},\text{rest}}$. Under the unbiasedness assumption $\mathbf{f}_{\mathbf{B},\text{rest}}$ is unbiased as well:

$$\mathrm{E}(\mathbf{f}_{\mathbf{B},\text{rest}}) = \mathbf{B}\boldsymbol{\mu}_{\mathbf{f}} = \sum_{i=1}^{k} \mathbf{B}_i \boldsymbol{\mu}_i = \sum_{i=1}^{k} \mathbf{B}_i \boldsymbol{\mu}_0 = \boldsymbol{\mu}_0 = \mathrm{E}(\mathbf{y}) \ . \tag{3.2}$$

Under the restriction $\sum_{i=1}^{k} \mathbf{B}_i = \mathbf{I}_l$ the error of the combined forecast results from the individual forecast errors $\mathbf{e}_i = \mathbf{f}_i - \mathbf{y}$ as a linear combination with exactly the same weights occurring in the combined forecast:

$$\begin{aligned} \mathbf{e}_{\mathbf{B},\text{rest}} &= \mathbf{f}_{\mathbf{B},\text{rest}} - \mathbf{y} = \sum_{i=1}^{k} \mathbf{B}_i \mathbf{f}_i - \left(\sum_{i=1}^{k} \mathbf{B}_i \right) \mathbf{y} \\ &= \sum_{i=1}^{k} \mathbf{B}_i (\mathbf{f}_i - \mathbf{y}) = \sum_{i=1}^{k} \mathbf{B}_i \mathbf{e}_i \ . \end{aligned} \tag{3.3}$$

Utilizing the restriction one more time we may continue

$$\begin{aligned} \mathbf{e}_{\mathbf{B},\text{rest}} &= \left(\mathbf{I}_l - \sum_{i=2}^{k} \mathbf{B}_i \right) \mathbf{e}_1 + \sum_{i=2}^{k} \mathbf{B}_i \mathbf{e}_i = \mathbf{e}_1 + \sum_{i=2}^{k} \mathbf{B}_i (\mathbf{e}_i - \mathbf{e}_1) \\ &= \mathbf{e}_1 + (\mathbf{B}_2 | \ldots | \mathbf{B}_k) \begin{pmatrix} \mathbf{e}_2 - \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_k - \mathbf{e}_1 \end{pmatrix} =: \mathbf{e}_1 + \mathbf{B}_{\text{red}} \mathbf{d} \ . \end{aligned} \tag{3.4}$$

Obviously, $\mathbf{B}_{\text{red}}$ is a reduced version of $\mathbf{B}$ from which the complete $\mathbf{B}$ can be reconstructed via the restriction. $\mathbf{d}$ is a $(k-1)l$-dimensional random vector, where we may also write $\mathbf{e}_i - \mathbf{e}_1 = (\mathbf{f}_i - \mathbf{y}) - (\mathbf{f}_1 - \mathbf{y}) = \mathbf{f}_i - \mathbf{f}_1$.
By the unbiasedness of the individual forecasts it follows that $\mathbf{e}_1$ is an unbiased statistic for the non-stochastic $\mathbf{0} \in \mathbb{R}^l$ and that $\mathbf{d}$ is unbiased for $\mathbf{0} \in \mathbb{R}^{(k-1)l}$. This is a situation where we can apply the (strong) covariance adjustment technique (Rao 1966, 1967), compare Lemma A.7 with $\mathbf{T} = \mathbf{e}_1$, $\boldsymbol{\theta} = \mathbf{0} \in \mathbb{R}^l$, $\mathbf{Z} = \mathbf{d}$ and $m = (k-1)l$. In order to obtain the optimal combination matrix $\mathbf{B}_{\text{red}}$ and hence $\mathbf{B}$ from Lemma A.7 we need to know the covariance matrix $\mathbf{W}$ of $(\mathbf{e}_1^{\mathrm{T}}, \mathbf{d}^{\mathrm{T}})^{\mathrm{T}}$. This covariance matrix may be calculated from the given second order moment matrix $\boldsymbol{\Sigma}$ belonging to the joint distribution of $\mathbf{y}$ and $\mathbf{f}$.
At this stage we have to make a short stop to evaluate the unbiasedness assumption connected with the combination $\mathbf{f}_{\mathbf{B},\text{rest}}$ and its effect on the second order moment matrix. Under the unbiasedness assumption we calculate the second order moment matrix as

$$\begin{aligned}
\breve{\boldsymbol{\Sigma}} &= \mathrm{E}\left[\left(\begin{pmatrix}\mathbf{y}\\ \mathbf{f}_1\\ \vdots\\ \mathbf{f}_k\end{pmatrix} - \begin{pmatrix}\boldsymbol{\mu}_0\\ \boldsymbol{\mu}_0\\ \vdots\\ \boldsymbol{\mu}_0\end{pmatrix}\right)\left(\begin{pmatrix}\mathbf{y}\\ \mathbf{f}_1\\ \vdots\\ \mathbf{f}_k\end{pmatrix} - \begin{pmatrix}\boldsymbol{\mu}_0\\ \boldsymbol{\mu}_0\\ \vdots\\ \boldsymbol{\mu}_0\end{pmatrix}\right)^{\mathrm{T}}\right] \\
&= \mathrm{E}\left[\left(\begin{pmatrix}\mathbf{y}\\ \mathbf{f}\end{pmatrix} - (\mathbf{1}_{k+1} \otimes \boldsymbol{\mu}_0)\right)\left(\begin{pmatrix}\mathbf{y}\\ \mathbf{f}\end{pmatrix} - (\mathbf{1}_{k+1} \otimes \boldsymbol{\mu}_0)\right)^{\mathrm{T}}\right], \qquad (3.5)
\end{aligned}$$

where $\otimes$ denotes the Kronecker product, i.e. $\mathbf{1}_{k+1} \otimes \boldsymbol{\mu}_0 = (\boldsymbol{\mu}_0^{\mathrm{T}}, \ldots, \boldsymbol{\mu}_0^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{(k+1)l}$. The unbiasedness property is assumed to hold when calculating the optimal combination parameters, but of course this assumption may in fact not be true. In this case the true second order moment matrix is given as

$$\begin{aligned}
\boldsymbol{\Sigma} &= \mathrm{E}\left[\left(\begin{pmatrix}\mathbf{y}\\ \mathbf{f}_1\\ \vdots\\ \mathbf{f}_k\end{pmatrix} - \begin{pmatrix}\boldsymbol{\mu}_0\\ \boldsymbol{\mu}_1\\ \vdots\\ \boldsymbol{\mu}_k\end{pmatrix}\right)\left(\begin{pmatrix}\mathbf{y}\\ \mathbf{f}_1\\ \vdots\\ \mathbf{f}_k\end{pmatrix} - \begin{pmatrix}\boldsymbol{\mu}_0\\ \boldsymbol{\mu}_1\\ \vdots\\ \boldsymbol{\mu}_k\end{pmatrix}\right)^{\mathrm{T}}\right] \\
&= \mathrm{E}\left[\left(\begin{pmatrix}\mathbf{y}\\ \mathbf{f}\end{pmatrix} - \begin{pmatrix}\boldsymbol{\mu}_0\\ \boldsymbol{\mu}_{\mathbf{f}}\end{pmatrix}\right)\left(\begin{pmatrix}\mathbf{y}\\ \mathbf{f}\end{pmatrix} - \begin{pmatrix}\boldsymbol{\mu}_0\\ \boldsymbol{\mu}_{\mathbf{f}}\end{pmatrix}\right)^{\mathrm{T}}\right]. \qquad (3.6)
\end{aligned}$$

Only if the unbiased assumption is correct the matrices $\boldsymbol{\Sigma}$ and $\breve{\boldsymbol{\Sigma}}$ coincide. Otherwise we can establish the following relationship:

$$\begin{aligned}
\breve{\boldsymbol{\Sigma}} &= \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} - \boldsymbol{\mu}(\mathbf{1}_{k+1} \otimes \boldsymbol{\mu}_0)^{\mathrm{T}} - (\mathbf{1}_{k+1} \otimes \boldsymbol{\mu}_0)\boldsymbol{\mu}^{\mathrm{T}} \\
&\quad + (\mathbf{1}_{k+1} \otimes \boldsymbol{\mu}_0)(\mathbf{1}_{k+1} \otimes \boldsymbol{\mu}_0)^{\mathrm{T}} \qquad (3.7)
\end{aligned}$$

and, consequently,

$$
\begin{aligned}
\breve{\mathbf{\Sigma}}_{00} &= \mathbf{\Sigma}_{00} \\
\breve{\mathbf{\Sigma}}_{\mathbf{f}0} &= \mathbf{\Sigma}_{\mathbf{f}0} \\
\breve{\mathbf{\Sigma}}_{\mathbf{ff}} &= \mathbf{\Sigma}_{\mathbf{ff}} + \boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}} - \boldsymbol{\mu}_{\mathbf{f}}(\mathbf{1}_k \otimes \boldsymbol{\mu}_0)^{\mathrm{T}} - (\mathbf{1}_k \otimes \boldsymbol{\mu}_0)\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}} \\
&\quad + (\mathbf{1}_k \otimes \boldsymbol{\mu}_0)(\mathbf{1}_k \otimes \boldsymbol{\mu}_0)^{\mathrm{T}} \, . \qquad (3.8)
\end{aligned}
$$

When considering the combination $\mathbf{f}_{\mathbf{B},\mathrm{rest}}$ we work with the unbiasedness assumption and hence we use the second order moment matrix $\breve{\mathbf{\Sigma}}$. We have to be aware, however, that the unbiasedness assumption may not be true and hence $\mathbf{\Sigma}$ and $\breve{\mathbf{\Sigma}}$ may differ.

Now we come back to the calculation of the covariance matrix $\mathbf{W}$ needed to apply the covariance adjustment technique. Of course, it will be calculated under the unbiasedness assumption, i.e. from $\breve{\mathbf{\Sigma}}$, and will, therefore, be denoted as $\breve{\mathbf{W}}$.

The current situation is the counterpart of univariate forecast combination with the combination weights adding up to one. In the corresponding literature the optimal combination weights are generally given in terms of the second order moments of the errors. To clarify the connection to our results we first want to relate the second order moments of the errors to $\breve{\mathbf{\Sigma}}$:

$$
\mathrm{Cov}\left[\begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_k \end{pmatrix}\right] =: \begin{pmatrix} \breve{\mathbf{V}}_{11} & \breve{\mathbf{V}}_{12} & \dots & \breve{\mathbf{V}}_{1k} \\ \breve{\mathbf{V}}_{21} & \breve{\mathbf{V}}_{22} & \dots & \breve{\mathbf{V}}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \breve{\mathbf{V}}_{k1} & \breve{\mathbf{V}}_{k2} & \dots & \breve{\mathbf{V}}_{kk} \end{pmatrix} = \breve{\mathbf{V}} \qquad (3.9)
$$

with

$$
\begin{aligned}
\breve{\mathbf{V}}_{ij} &= \mathrm{Cov}(\mathbf{e}_i, \mathbf{e}_j) = \mathrm{Cov}(\mathbf{f}_i - \mathbf{y}, \mathbf{f}_j - \mathbf{y}) \\
&= \mathrm{Cov}(\mathbf{f}_i, \mathbf{f}_j) - \mathrm{Cov}(\mathbf{f}_i, \mathbf{y}) - \mathrm{Cov}(\mathbf{y}, \mathbf{f}_j) + \mathrm{Cov}(\mathbf{y}, \mathbf{y}) \\
&= \breve{\mathbf{\Sigma}}_{ij} - \breve{\mathbf{\Sigma}}_{i0} - \breve{\mathbf{\Sigma}}_{0j} + \breve{\mathbf{\Sigma}}_{00} \, . \qquad (3.10)
\end{aligned}
$$

From this we further calculate

$$
\mathrm{Cov}\left[\begin{pmatrix} \mathbf{e}_1 \\ \mathbf{d} \end{pmatrix}\right] =: \begin{pmatrix} \breve{\mathbf{W}}_{11} & \breve{\mathbf{W}}_{12} \\ \breve{\mathbf{W}}_{12}^{\mathrm{T}} & \breve{\mathbf{W}}_{22} \end{pmatrix} = \breve{\mathbf{W}} \, , \qquad (3.11)
$$

where

$$\breve{\mathbf{W}}_{11} = \mathrm{Cov}(\mathbf{e}_1) = \breve{\mathbf{V}}_{11} = \breve{\boldsymbol{\Sigma}}_{11} - \breve{\boldsymbol{\Sigma}}_{10} - \breve{\boldsymbol{\Sigma}}_{01} + \breve{\boldsymbol{\Sigma}}_{00} \ , \tag{3.12}$$

$$\begin{aligned} \breve{\mathbf{W}}_{12} &= \mathrm{Cov}(\mathbf{e}_1, \mathbf{d}) = \mathrm{Cov}\left[\mathbf{e}_1, \begin{pmatrix} \mathbf{e}_2 - \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_k - \mathbf{e}_1 \end{pmatrix}\right] \\ &= (\breve{\mathbf{V}}_{12} - \breve{\mathbf{V}}_{11}, \dots, \breve{\mathbf{V}}_{1k} - \breve{\mathbf{V}}_{11}) \\ &= (\breve{\boldsymbol{\Sigma}}_{12} - \breve{\boldsymbol{\Sigma}}_{11} - \breve{\boldsymbol{\Sigma}}_{02} + \breve{\boldsymbol{\Sigma}}_{01}, \dots, \breve{\boldsymbol{\Sigma}}_{1k} - \breve{\boldsymbol{\Sigma}}_{11} - \breve{\boldsymbol{\Sigma}}_{0k} + \breve{\boldsymbol{\Sigma}}_{01}) \quad \text{and} \end{aligned} \tag{3.13}$$

$$\begin{aligned} \breve{\mathbf{W}}_{22} &= \mathrm{Cov}(\mathbf{d}) = \mathrm{Cov}\left[\begin{pmatrix} \mathbf{e}_2 - \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_k - \mathbf{e}_1 \end{pmatrix}, \begin{pmatrix} \mathbf{e}_2 - \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_k - \mathbf{e}_1 \end{pmatrix}\right] \\ &= (\mathrm{Cov}(\mathbf{e}_i - \mathbf{e}_1, \mathbf{e}_s - \mathbf{e}_1))_{i,s=2,\dots,k} \end{aligned} \tag{3.14}$$

with

$$\begin{aligned} \mathrm{Cov}(\mathbf{e}_i - \mathbf{e}_1, \mathbf{e}_s - \mathbf{e}_1) &= \breve{\mathbf{V}}_{is} - \breve{\mathbf{V}}_{i1} - \breve{\mathbf{V}}_{1s} + \breve{\mathbf{V}}_{11} \\ &= \breve{\boldsymbol{\Sigma}}_{is} - \breve{\boldsymbol{\Sigma}}_{i1} - \breve{\boldsymbol{\Sigma}}_{1s} + \breve{\boldsymbol{\Sigma}}_{11} \ . \end{aligned} \tag{3.15}$$

Now knowing $\breve{\mathbf{W}}$ we may derive the optimal combination parameters for $\mathbf{f}_{\mathbf{B},\mathrm{rest}}$ from Lemma A.7:

$$\mathbf{B}_{\mathrm{red,opt}} = (\mathbf{B}_{2,\mathrm{opt}}|\dots|\mathbf{B}_{k,\mathrm{opt}}) = -\breve{\mathbf{W}}_{12}\breve{\mathbf{W}}_{22}^{-1} \quad \text{and} \quad \mathbf{B}_{1,\mathrm{opt}} = \mathbf{I}_l - \sum_{i=2}^{k} \mathbf{B}_{i,\mathrm{opt}} \ . \tag{3.16}$$

In deriving the optimal value of the SMSPE-function corresponding to the above optimal combination parameters we have to distinguish between two situations: If the individual forecasts $\mathbf{f}_i$ and hence also $\mathbf{f}_{\mathbf{B},\mathrm{rest}}$ are actually unbiased, we have $\mathrm{E}(\mathbf{e}_{\mathbf{B},\mathrm{rest}}) = \mathbf{0}$ and, thus,

$$\mathrm{Cov}(\mathbf{e}_{\mathbf{B},\mathrm{rest}}) = \mathrm{E}[(\mathbf{e}_{\mathbf{B},\mathrm{rest}})(\mathbf{e}_{\mathbf{B},\mathrm{rest}})^{\mathrm{T}}] = \mathrm{MMSPE}(\mathbf{f}_{\mathbf{B},\mathrm{rest}}, \mathbf{y}) \ . \tag{3.17}$$

Consequently, from Lemma A.7 we obtain the optimal value of the MMSPE-function as

$$\mathrm{MMSPE}(\mathbf{f}_{\mathbf{B}_{\mathrm{opt}},\mathrm{rest}}, \mathbf{y}) = \breve{\mathbf{W}}_{11} - \breve{\mathbf{W}}_{12}\breve{\mathbf{W}}_{22}^{-1}\breve{\mathbf{W}}_{21} \ , \tag{3.18}$$

from which the corresponding optimal value of the SMSPE-function results by taking traces

$$\mathrm{SMSPE}(\mathbf{f}_{\mathbf{B}_{\mathrm{opt}},\mathrm{rest}}, \mathbf{y}) = \mathrm{tr}(\breve{\mathbf{W}}_{11}) - \mathrm{tr}(\breve{\mathbf{W}}_{12}\breve{\mathbf{W}}_{22}^{-1}\breve{\mathbf{W}}_{21}) \ . \tag{3.19}$$

If, however, we cannot assume unbiasedness we have to calculate the optimal SMSPE-value by inserting the derived optimal combination parameters into the general SMSPE-function given in Equation (2.5).
Since the (strong) covariance adjustment technique minimizes the covariance *matrix* criterion it is clear that the above choice of the parameter matrix $\mathbf{B}$ is optimal with respect to both criteria, MMSPE and SMSPE.

Now we turn to the combined forecast

$$\mathbf{f}_{\mathbf{B},\mathbf{c},\mathrm{rest}} = \mathbf{B}\mathbf{f} + \mathbf{c} = \sum_{i=1}^{k} \mathbf{B}_i \mathbf{f}_i + \mathbf{c} \quad \text{with} \quad \sum_{i=1}^{k} \mathbf{B}_i = \mathbf{I}_l \ . \tag{3.20}$$

Obviously, we no longer assume the individual forecasts to be unbiased, otherwise there would be no need for introducing the constant term $\mathbf{c}$.
The constant term can only influence the bias of the combined forecast but not its covariance matrix. Therefore the optimal choice for $\mathbf{c}$ is necessarily the one that makes the combined forecast $\mathbf{f}_{\mathbf{B}_{\mathrm{opt}},\mathbf{c}_{\mathrm{opt}},\mathrm{rest}}$ unbiased, i.e.

$$\mathbf{c}_{\mathrm{opt}} = \boldsymbol{\mu}_0 - \mathbf{B}_{\mathrm{opt}} \boldsymbol{\mu}_{\mathbf{f}} \ . \tag{3.21}$$

To find the optimal choice for $\mathbf{B}$ we may exploit this fact and consider the SMSPE of

$$\begin{aligned} \mathbf{B}\mathbf{f} + \boldsymbol{\mu}_0 - \mathbf{B}\boldsymbol{\mu}_{\mathbf{f}} &= \sum_{i=1}^{k} \mathbf{B}_i \mathbf{f}_i + \left(\sum_{i=1}^{k} \mathbf{B}_i \right) \boldsymbol{\mu}_0 - \sum_{i=1}^{k} \mathbf{B}_i \boldsymbol{\mu}_i \\ &= \sum_{i=1}^{k} \mathbf{B}_i (\mathbf{f}_i - \boldsymbol{\mu}_i + \boldsymbol{\mu}_0) \ , \end{aligned} \tag{3.22}$$

where the restriction $\sum_{i=1}^{k} \mathbf{B}_i = \mathbf{I}_l$ has been used as well.
Since the $\mathbf{f}_i - \boldsymbol{\mu}_i + \boldsymbol{\mu}_0$ are unbiased we are back in the previous case, where we wanted to find the optimal weights $\mathbf{B}_i$ for a linear combination of unbiased forecasts without constant term but under the restriction of the combination weights summing up to the identity matrix. Since the $\mathbf{f}_i - \boldsymbol{\mu}_i + \boldsymbol{\mu}_0$ result from shifting the $\mathbf{f}_i$ by a constant vector, their second order moments are the same. Consequently, the optimal choice for $\mathbf{B}$ is given by

$$\mathbf{B}_{\mathrm{red,opt}} = (\mathbf{B}_{2,\mathrm{opt}} | \ldots | \mathbf{B}_{k,\mathrm{opt}}) = -\mathbf{W}_{12}\mathbf{W}_{22}^{-1} \quad \text{and} \quad \mathbf{B}_{1,\mathrm{opt}} = \mathbf{I}_l - \sum_{i=2}^{k} \mathbf{B}_{i,\mathrm{opt}} \ , \tag{3.23}$$

and the corresponding optimal SMSPE-value is given by

$$\mathrm{SMSPE}(\mathbf{f}_{\mathbf{B}_{\mathrm{opt}},\mathbf{c}_{\mathrm{opt}},\mathrm{rest}},\mathbf{y}) = \mathrm{tr}(\mathbf{W}_{11}) - \mathrm{tr}(\mathbf{W}_{12}\mathbf{W}_{22}^{-1}\mathbf{W}_{21}) \ . \tag{3.24}$$

There is, however, no unbiasedness assumption in this case and for this reason we now have to use the second order moment matrices $\mathbf{\Sigma}$, $\mathbf{V}$ and $\mathbf{W}$ instead of $\breve{\mathbf{\Sigma}}$, $\breve{\mathbf{V}}$ and $\breve{\mathbf{W}}$. $\mathbf{V}$ and $\mathbf{W}$ are calculated from $\mathbf{\Sigma}$ in exactly the same way as $\breve{\mathbf{V}}$ and $\breve{\mathbf{W}}$ are calculated from $\breve{\mathbf{\Sigma}}$.
Similarly to the previous case it is clear that we have SMSPE- as well as MMSPE-optimality of the chosen combination parameters.

When applying the combinations to empirical data one often has to face the difficulty that the moments $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are unknown and, thus, have to be estimated from the data. In this context it is important that the estimates are chosen appropriately, i.e. for $\mathbf{f}_{\mathbf{B}_{\mathrm{opt}},\mathrm{rest}}$ we have to estimate with the unbiasedness assumption and for $\mathbf{f}_{\mathbf{B}_{\mathrm{opt}},\mathbf{c}_{\mathrm{opt}},\mathrm{rest}}$ we have to estimate without this assumption. Compare also THIELE (1993, Sections 3.1.2 and 3.3.2) for a corresponding result in the univariate setting. The question arises which is the proper method to estimate $\breve{\mathbf{\Sigma}}$ and hence $\mathbf{B}_{\mathrm{opt}}$ under the unbiasedness assumption. How should the joint expectation $\boldsymbol{\mu}_0$ of $\mathbf{y}$ and the forecasts $\mathbf{f}_1, \ldots, \mathbf{f}_k$ be estimated? As the mean of the observations on $\mathbf{y}$, as the mean of all observations on $\mathbf{y}$ and the $\mathbf{f}_i$, or otherwise? The following lemma implies that the estimate for $\mathbf{B}_{\mathrm{opt}}$ is insensitive with respect to the estimation of $\boldsymbol{\mu}_0$, we may even use the constant $\mathbf{0}$. The lemma is proven in Appendix B.

Lemma 3.1 *The optimal parameter matrix* $\mathbf{B}_{\mathrm{opt}}$ *for the combination* $\mathbf{f}_{\mathbf{B}_{\mathrm{opt}},\mathrm{rest}}$ *is not altered if we use any other constant vector than* $\boldsymbol{\mu}_0$ *in the calculation of the covariance matrix*

$$\breve{\mathbf{\Sigma}} = \mathrm{E}\left[\left(\begin{pmatrix}\mathbf{y}\\ \mathbf{f}\end{pmatrix} - (\mathbf{1}_{k+1}\otimes\boldsymbol{\mu}_0)\right)\left(\begin{pmatrix}\mathbf{y}\\ \mathbf{f}\end{pmatrix} - (\mathbf{1}_{k+1}\otimes\boldsymbol{\mu}_0)\right)^{\mathrm{T}}\right] .$$

In Section 4 we will introduce variants of the linear combinations requiring the estimation of less parameters. They may prove useful especially if there are only few data to estimate the combination parameters.

4 Variants involving less parameters

Considering the multivariate combination approaches introduced in Section 2 it becomes obvious that they utilize a large number of parameters: $\mathbf{f}_{\mathbf{B},\mathbf{c}}$ involves a

total of $l^2k + l$ parameters, $\mathbf{f_B}$ depends on l^2k, $\mathbf{f}_{\mathbf{B},\text{rest}}$ on $l^2(k-1)$ and $\mathbf{f}_{\mathbf{B},\mathbf{c},\text{rest}}$ on $l^2(k-1)+l$ parameters. Even for relatively few forecasts $\mathbf{f}_1, \ldots, \mathbf{f}_k$ and a relatively small dimension l of the vector $\mathbf{y}$ the number of parameters becomes quite large.
Therefore, it seems worthwhile to look for variants of the considered linear approaches involving less parameters but still capturing the spirit of multivariate combination. Our approach is to restrict the parameter matrix $\mathbf{B} = (\mathbf{B}_1 \vdots \ldots \vdots \mathbf{B}_k) \in \mathbb{R}^{l \times kl}$ from the linear approaches such that the number of parameters is reduced.
If we restrict the full $l \times l$ weight matrices $\mathbf{B}_i$ to be diagonal matrices

$$\mathbf{D}_i = \begin{pmatrix} D_{i,11} & 0 & \ldots & 0 \\ 0 & D_{i,22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & D_{i,ll} \end{pmatrix} \tag{4.1}$$

we reduce the number of parameters in each weight matrix from l^2 to l. The resulting linear combination without further restrictions on the $\mathbf{D}_i$ and containing a constant term $\mathbf{c}$, for example, is

$$\mathbf{f}_{\mathbf{D},\mathbf{c}} = \mathbf{Df} + \mathbf{c} = \mathbf{D}_1\mathbf{f}_1 + \ldots + \mathbf{D}_k\mathbf{f}_k + \mathbf{c} \ , \tag{4.2}$$

where $\mathbf{D} = (\mathbf{D}_1 | \ldots | \mathbf{D}_k)$. Now only $l(k+1)$ parameters are involved. The other three combinations $\mathbf{f_D}$, $\mathbf{f}_{\mathbf{D},\text{rest}}$ and $\mathbf{f}_{\mathbf{D},\mathbf{c},\text{rest}}$ are defined analogously.
If we further restrict the weight matrices $\mathbf{B}_i$ to be scalar multiples $\alpha_i \mathbf{I}_l$ of the $l \times l$ identity matrix, we further reduce the number of parameters in each weight matrix to 1. The resulting linear combination without further restrictions on the α_i and having a constant term $\mathbf{c}$, for example, is

$$\mathbf{f}_{\boldsymbol{\alpha},\mathbf{c}} = \alpha_1\mathbf{f}_1 + \ldots + \alpha_k\mathbf{f}_k + \mathbf{c} \ , \tag{4.3}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)^{\mathrm{T}}$ and only $k + l$ parameters are involved. The other three combinations $\mathbf{f}_{\boldsymbol{\alpha}}$, $\mathbf{f}_{\boldsymbol{\alpha},\text{rest}}$ and $\mathbf{f}_{\boldsymbol{\alpha},\mathbf{c},\text{rest}}$ are defined analogously.
We will refer to the combinations with full weight matrices $\mathbf{B}_i$ as *strong* combinations, to the combinations with diagonal weight matrices $\mathbf{D}_i$ as *medium* combinations and to the combinations with scalar weight matrices $\alpha_i \mathbf{I}_l$ as *weak* combinations.
We will see that the weak combinations under the restriction $\sum_{i=1}^k \alpha_i = 1$ of the combination weights adding up to one can be interpreted in terms of covariance adjustment as well, more precisely in terms of *weak* covariance adjustment as introduced by Trenkler and Ihorst (1995). This correspondence is another reason to look at weak combinations.

4.1 Medium combinations

Inspecting the medium combination approach more closely it becomes obvious that each component of the forecasts is combined separately: The forecasts $f_{1,j}, \ldots, f_{k,j}$ for the j-th component y_j of the target vector $\mathbf{y}$ are combined in the fashion of univariate forecast combination using the parameters $D_{1,jj}, \ldots, D_{k,jj}$ and c_j. Consequently, the minimum SMSPE combination of the forecast vectors $\mathbf{f}_i$ within the medium setting is obtained by choosing $D_{1,jj}, \ldots, D_{k,jj}$ and c_j according to the univariate MSPE-optimal choices for each $j = 1, \ldots, l$. These choices are listed in Troschke and Trenkler (2000). Alternatively, they may be obtained from the strong multivariate linear combinations in Sections 2 and 3 as the special case of $l = 1$ dimensional target variable and forecasts. In part these results can also be found in Harville (1985).
We will consider the medium combination with constant term and no restriction on the sum of the $\mathbf{D}_i$ as an example, i.e. we consider $\mathbf{f}_{\mathbf{D},\mathbf{c}} = \mathbf{D}_1\mathbf{f}_1 + \ldots + \mathbf{D}_k\mathbf{f}_k + \mathbf{c}$ with expectation

$$\mathrm{E}(\mathbf{f}_{\mathbf{D},\mathbf{c}}) = \mathbf{D}_1\boldsymbol{\mu}_1 + \ldots + \mathbf{D}_k\boldsymbol{\mu}_k + \mathbf{c}\ . \tag{4.4}$$

Using the abbreviations $\tilde{\mathbf{f}}_j = (f_{1,j}, \ldots, f_{k,j})^{\mathrm{T}}$, $\tilde{\boldsymbol{\mu}}_j = (\mu_{1,j}, \ldots, \mu_{k,j})^{\mathrm{T}}$ and $\tilde{\mathbf{D}}_j = (D_{1,jj}, \ldots, D_{k,jj})^{\mathrm{T}}$ we conclude from the results in the univariate case that the choices for c_j and $\tilde{\mathbf{D}}_j$ minimizing the univariate MSPE-criterion are given by

$$(\tilde{\mathbf{D}}_j)_{\mathrm{opt}} = \boldsymbol{\Sigma}_{\tilde{\mathbf{f}}_j\tilde{\mathbf{f}}_j}^{-1}\,\mathrm{Cov}(\tilde{\mathbf{f}}_j, y_j) \quad \text{and} \quad (c_j)_{\mathrm{opt}} = \mu_{0,j} - \mathrm{Cov}(y_j, \tilde{\mathbf{f}}_j)\boldsymbol{\Sigma}_{\tilde{\mathbf{f}}_j\tilde{\mathbf{f}}_j}^{-1}\tilde{\boldsymbol{\mu}}_j\ , \tag{4.5}$$

where $\boldsymbol{\Sigma}_{\tilde{\mathbf{f}}_j\tilde{\mathbf{f}}_j} = \mathrm{Cov}(\tilde{\mathbf{f}}_j, \tilde{\mathbf{f}}_j)$. Invertibility of $\boldsymbol{\Sigma}_{\tilde{\mathbf{f}}_j\tilde{\mathbf{f}}_j}$ is granted because $\boldsymbol{\Sigma}_{\mathbf{ff}}$ is positive definite, cf. Lemma A.9 with $\mathbf{A} = \boldsymbol{\Sigma}_{\mathbf{ff}}$ and $\mathbf{A}_j = \boldsymbol{\Sigma}_{\tilde{\mathbf{f}}_j\tilde{\mathbf{f}}_j}$. By these $(\tilde{\mathbf{D}}_j)_{\mathrm{opt}}$ and $(c_j)_{\mathrm{opt}}$ for $j = 1, \ldots, l$ the optimal choices $\mathbf{D}_{\mathrm{opt}}$ and $\mathbf{c}_{\mathrm{opt}}$ are determined.
Since

$$\begin{aligned}
\mathrm{SMSPE}(\mathbf{f}_{\mathbf{D},\mathbf{c}}, \mathbf{y}) &= \mathrm{E}[(\mathbf{y} - \mathbf{f}_{\mathbf{D},\mathbf{c}})^{\mathrm{T}}(\mathbf{y} - \mathbf{f}_{\mathbf{D},\mathbf{c}})] \\
&= \mathrm{E}[\sum_{j=1}^{l}(y_j - (\mathbf{f}_{\mathbf{D},\mathbf{c}})_j)^2] \\
&= \sum_{j=1}^{l}\mathrm{E}[(y_j - (\mathbf{f}_{\mathbf{D},\mathbf{c}})_j)^2] \\
&= \sum_{j=1}^{l}\mathrm{MSPE}((\mathbf{f}_{\mathbf{D},\mathbf{c}})_j, y_j)
\end{aligned} \tag{4.6}$$

and since each component is treated separately by the medium approach, the optimal SMSPE-value is obtained as the sum of the optimal MSPE-values in each component, i.e.

$$\begin{aligned}
\mathrm{SMSPE}(\mathbf{f}_{\mathbf{D}_{\mathrm{opt}},\mathbf{c}_{\mathrm{opt}}},\mathbf{y}) &= \sum_{j=1}^{l} \mathrm{MSPE}(\mathbf{f}_{(\tilde{\mathbf{D}}_j)_{\mathrm{opt}},(\mathbf{c}_j)_{\mathrm{opt}}}, y_j) \\
&= \sum_{j=1}^{l} \left[\Sigma_{00,jj} - \mathrm{Cov}(y_j,\tilde{\mathbf{f}}_j)\boldsymbol{\Sigma}^{-1}_{\tilde{\mathbf{f}}_j\tilde{\mathbf{f}}_j}\,\mathrm{Cov}(\tilde{\mathbf{f}}_j,y_j)\right] \\
&= \mathrm{tr}(\boldsymbol{\Sigma}_{00}) - \sum_{j=1}^{l} \mathrm{Cov}(y_j,\tilde{\mathbf{f}}_j)\boldsymbol{\Sigma}^{-1}_{\tilde{\mathbf{f}}_j\tilde{\mathbf{f}}_j}\,\mathrm{Cov}(\tilde{\mathbf{f}}_j,y_j)\ , \qquad (4.7)
\end{aligned}$$

where $\Sigma_{00,jj}$ denotes the j-th diagonal element of $\boldsymbol{\Sigma}_{00}$.
To obtain these results as the special case $l=1$ from the strong combination we set $\mathbf{f}=\tilde{\mathbf{f}}_j$, $\mathbf{B}=\tilde{\mathbf{D}}_j^{\mathrm{T}}$, $\mathbf{c}=c_j$, $\boldsymbol{\mu}_0=\mu_{0,j}$, $\boldsymbol{\mu}_{\mathbf{f}}=\tilde{\boldsymbol{\mu}}_j$, $\boldsymbol{\Sigma}_{00}=\Sigma_{00,jj}$, $\boldsymbol{\Sigma}_{0\mathbf{f}}=\mathrm{Cov}(y_j,\tilde{\mathbf{f}}_j)$ and $\boldsymbol{\Sigma}_{\mathbf{ff}}=\boldsymbol{\Sigma}_{\tilde{\mathbf{f}}_j\tilde{\mathbf{f}}_j}$ in the formulae for the combination $\mathbf{f}_{\mathbf{B},\mathbf{c}}$.

4.2 Weak combinations

We will now derive the optimal combination parameters within the variants of the weak approach. First let us consider the weak combination with constant term and with no restriction on the sum of the α_i, i.e. we consider $\mathbf{f}_{\boldsymbol{\alpha},\mathbf{c}}=\alpha_1\mathbf{f}_1+\ldots+\alpha_k\mathbf{f}_k+\mathbf{c}$ with expectation

$$\mathrm{E}(\mathbf{f}_{\boldsymbol{\alpha},\mathbf{c}}) = \alpha_1\boldsymbol{\mu}_1+\ldots+\alpha_k\boldsymbol{\mu}_k+\mathbf{c}\ . \qquad (4.8)$$

Again, we will have to explicitly calculate the SMSPE-function of $\mathbf{f}_{\boldsymbol{\alpha},\mathbf{c}}$, differentiate this function with respect to the occurring parameters $\boldsymbol{\alpha}$ and $\mathbf{c}$, and finally equate the derivatives simultaneously to zero.

Step 1: Explicit calculation of the SMSPE**-function.** By inserting $\mathbf{B}=(\alpha_1\mathbf{I}_l|\ldots|\alpha_k\mathbf{I}_l)$ into the general SMSPE-formula (2.5), utilizing our notation from Section 1 and finally applying Lemma A.2 we find the scalar mean square prediction error of $\mathbf{f}_{\boldsymbol{\alpha},\mathbf{c}}$ as follows:

$$
\begin{aligned}
\mathrm{SMSPE}(\mathbf{f}_{\boldsymbol{\alpha},\mathbf{c}}, \mathbf{y}) = \\
&\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{H}\boldsymbol{\alpha} \\
&- 2\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{h} \\
&+ 2\mathbf{c}^{\mathrm{T}}\left(\sum_{i=1}^{k}\alpha_i\boldsymbol{\mu}_i\right) \\
&+ \mathbf{c}^{\mathrm{T}}\mathbf{c} \\
&- 2\boldsymbol{\mu}_0^{\mathrm{T}}\mathbf{c} \\
&+ \mathrm{tr}(\boldsymbol{\Sigma}_{00}) + \boldsymbol{\mu}_0^{\mathrm{T}}\boldsymbol{\mu}_0 \ ,
\end{aligned}
\tag{4.9}
$$

where

$$
\begin{aligned}
\mathbf{H} &:= \begin{pmatrix} \mathrm{tr}(\boldsymbol{\Sigma}_{11}) + \boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\mu}_1 & \dots & \mathrm{tr}(\boldsymbol{\Sigma}_{1k}) + \boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\mu}_k \\ \vdots & & \vdots \\ \mathrm{tr}(\boldsymbol{\Sigma}_{k1}) + \boldsymbol{\mu}_k^{\mathrm{T}}\boldsymbol{\mu}_1 & \dots & \mathrm{tr}(\boldsymbol{\Sigma}_{kk}) + \boldsymbol{\mu}_k^{\mathrm{T}}\boldsymbol{\mu}_k \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{tr}(\boldsymbol{\Sigma}_{11}) & \dots & \mathrm{tr}(\boldsymbol{\Sigma}_{1k}) \\ \vdots & & \vdots \\ \mathrm{tr}(\boldsymbol{\Sigma}_{k1}) & \dots & \mathrm{tr}(\boldsymbol{\Sigma}_{kk}) \end{pmatrix} + \begin{pmatrix} \boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\mu}_1 & \dots & \boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\mu}_k \\ \vdots & & \vdots \\ \boldsymbol{\mu}_k^{\mathrm{T}}\boldsymbol{\mu}_1 & \dots & \boldsymbol{\mu}_k^{\mathrm{T}}\boldsymbol{\mu}_k \end{pmatrix} \\
&=: \mathbf{H}_1 + \mathbf{H}_2
\end{aligned}
\tag{4.10}
$$

are $(k \times k)$ symmetric matrices and

$$
\mathbf{h} := \begin{pmatrix} \mathrm{tr}(\boldsymbol{\Sigma}_{10}) + \boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\mu}_0 \\ \vdots \\ \mathrm{tr}(\boldsymbol{\Sigma}_{k0}) + \boldsymbol{\mu}_k^{\mathrm{T}}\boldsymbol{\mu}_0 \end{pmatrix} = \begin{pmatrix} \mathrm{tr}(\boldsymbol{\Sigma}_{10}) \\ \vdots \\ \mathrm{tr}(\boldsymbol{\Sigma}_{k0}) \end{pmatrix} + \begin{pmatrix} \boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\mu}_0 \\ \vdots \\ \boldsymbol{\mu}_k^{\mathrm{T}}\boldsymbol{\mu}_0 \end{pmatrix} =: \mathbf{h}_1 + \mathbf{h}_2 \tag{4.11}
$$

are k-dimensional vectors.

Note that by applying Lemma A.2 we may rewrite the third term of the above function as

$$
2\mathbf{c}^{\mathrm{T}}\left(\sum_{i=1}^{k}\alpha_i\boldsymbol{\mu}_i\right) = 2\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{q}_{\mathbf{c}} \ , \tag{4.12}
$$

where

$$
\mathbf{q}_{\mathbf{c}} = \begin{pmatrix} \mathbf{c}^{\mathrm{T}}\boldsymbol{\mu}_1 \\ \vdots \\ \mathbf{c}^{\mathrm{T}}\boldsymbol{\mu}_k \end{pmatrix} \ , \tag{4.13}
$$

which is useful for the subsequent calculation of the derivatives.

Step 2: Differentiation. Applying Lemma (A.5) we get

$$\frac{\partial \mathrm{SMSPE}(\mathbf{f}_{\boldsymbol{\alpha},\mathbf{c}}, \mathbf{y})}{\partial \mathbf{c}} = 2\left[\mathbf{c} - \boldsymbol{\mu}_0 + \sum_{i=1}^{k} \alpha_i \boldsymbol{\mu}_i\right] \ . \tag{4.14}$$

and

$$\frac{\partial \mathrm{SMSPE}(\mathbf{f}_{\boldsymbol{\alpha},\mathbf{c}}, \mathbf{y})}{\partial \boldsymbol{\alpha}} = 2\left[\mathbf{H}\boldsymbol{\alpha} - \mathbf{h} + \mathbf{q}_{\mathbf{c}}\right] \ . \tag{4.15}$$

Step 3: Equating to zero. Setting Equations (4.14) and (4.15) simultaneously to zero and solving the resulting linear equation system for the unknown parameters we obtain the optimal choices for $\boldsymbol{\alpha}$ and $\mathbf{c}$.
From Equation (4.14) we get

$$\mathbf{c}_{\mathrm{opt}} = \boldsymbol{\mu}_0 - \sum_{i=1}^{k} \alpha_{i,\mathrm{opt}} \boldsymbol{\mu}_i \ . \tag{4.16}$$

Setting the derivative (4.15) equal to zero and using the result (4.16) after some lengthy calculations (involving reformulations like in Equations (4.12) and (4.13)) we finally obtain $\mathbf{H}_1 \boldsymbol{\alpha}_{\mathrm{opt}} = \mathbf{h}_1$ or equivalently

$$\boldsymbol{\alpha}_{\mathrm{opt}} = \mathbf{H}_1^{-1} \mathbf{h}_1 \ . \tag{4.17}$$

Invertibility of $\mathbf{H}_1$ is granted by the fact that $\boldsymbol{\Sigma}_{\mathbf{ff}}$ is positive definite, cf. Lemma A.9 with $\mathbf{A} = \boldsymbol{\Sigma}_{\mathbf{ff}}$ and $\mathbf{A}_{\mathrm{tr}} = \mathbf{H}_1$.
By the same reasoning as in the strong case it is obvious that this solution minimizes the SMSPE-function within the considered class of weakly combined forecasts, cf. Equation (2.11). From Equation (4.8) we may conclude that the combined forecast $\mathbf{f}_{\boldsymbol{\alpha}_{\mathrm{opt}},\mathbf{c}_{\mathrm{opt}}}$ is unbiased even if the single forecasts are biased.
Inserting $(\boldsymbol{\alpha}_{\mathrm{opt}}, \mathbf{c}_{\mathrm{opt}})$ into Equation (4.9) we may derive that the corresponding optimal SMSPE-value is given by

$$\mathrm{SMSPE}(\mathbf{f}_{\boldsymbol{\alpha}_{\mathrm{opt}},\mathbf{c}_{\mathrm{opt}}}, \mathbf{y}) = \mathrm{tr}(\boldsymbol{\Sigma}_{00}) - \mathbf{h}_1^{\mathrm{T}} \mathbf{H}_1^{-1} \mathbf{h}_1 \ . \tag{4.18}$$

See Section 6 for an analysis of the potential offered by this combination in a simple example.

The weak combination without constant term and without restriction on the sum of the α_i is

$$\mathbf{f}_{\boldsymbol{\alpha}} = \alpha_1 \mathbf{f}_1 + \ldots + \alpha_k \mathbf{f}_k \tag{4.19}$$

with expectation

$$\mathrm{E}(\mathbf{f}_{\boldsymbol{\alpha}}) = \alpha_1 \boldsymbol{\mu}_1 + \ldots + \alpha_k \boldsymbol{\mu}_k \ . \tag{4.20}$$

The corresponding SMSPE-function is obtained by setting $\mathbf{c} = \mathbf{0}$ in Equation (4.9):

$$\mathrm{SMSPE}(\mathbf{f}_{\boldsymbol{\alpha}}, \mathbf{y}) = \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{H} \boldsymbol{\alpha} - 2\boldsymbol{\alpha}^{\mathrm{T}} \mathbf{h} + \mathrm{tr}(\boldsymbol{\Sigma}_{00}) + \boldsymbol{\mu}_0^{\mathrm{T}} \boldsymbol{\mu}_0 \ . \tag{4.21}$$

This expression is minimized by

$$\boldsymbol{\alpha}_{\mathrm{opt}} = \mathbf{H}^{-1} \mathbf{h} \ , \tag{4.22}$$

where invertibility of $\mathbf{H}$ is granted by the positive definiteness of $\boldsymbol{\Sigma}_{\mathbf{ff}} + \boldsymbol{\mu}_{\mathbf{f}} \boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}$ and Lemma A.9 with $\mathbf{A} = \boldsymbol{\Sigma}_{\mathbf{ff}} + \boldsymbol{\mu}_{\mathbf{f}} \boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}$ and $\mathbf{A}_{\mathrm{tr}} = \mathbf{H}$. Obviously, $\mathbf{f}_{\boldsymbol{\alpha}_{\mathrm{opt}}}$ need not be unbiased even if the single forecasts $\mathbf{f}_i$ are unbiased.
Inserting $\boldsymbol{\alpha}_{\mathrm{opt}}$ into Equation (4.21) leads to the optimal value of the SMSPE-function within the considered class of combinations:

$$\mathrm{SMSPE}(\mathbf{f}_{\boldsymbol{\alpha}_{\mathrm{opt}}}, \mathbf{y}) = \mathrm{tr}(\boldsymbol{\Sigma}_{00}) + \boldsymbol{\mu}_0^{\mathrm{T}} \boldsymbol{\mu}_0 - \mathbf{h}^{\mathrm{T}} \mathbf{H}^{-1} \mathbf{h} \ . \tag{4.23}$$

The remaining two weak combinations $\mathbf{f}_{\boldsymbol{\alpha},\mathrm{rest}}$ and $\mathbf{f}_{\boldsymbol{\alpha},\mathbf{c},\mathrm{rest}}$ utilize the restriction $\mathbf{1}_k^{\mathrm{T}} \boldsymbol{\alpha} = 1$ of the scalar weights summing up to one. Like in the case of full weight matrices (cf. Section 3) it is advantageous to choose a representation via the forecast errors. Then the weak covariance adjustment technique (Ihorst, 1993, Trenkler and Ihorst, 1995) can be applied.
Like in the case of $\mathbf{f}_{\mathbf{B},\mathrm{rest}}$ for the combined forecast

$$\mathbf{f}_{\boldsymbol{\alpha},\mathrm{rest}} = (\boldsymbol{\alpha}^{\mathrm{T}} \otimes \mathbf{I}_l)\mathbf{f} = \sum_{i=1}^{k} \alpha_i \mathbf{f}_i \quad \text{with} \quad \sum_{i=1}^{k} \alpha_i = 1 \tag{4.24}$$

we assume unbiasedness of each single forecast $\mathbf{f}_i$. Then also $\mathbf{f}_{\boldsymbol{\alpha},\mathrm{rest}}$ is unbiased:

$$\mathrm{E}(\mathbf{f}_{\boldsymbol{\alpha},\mathrm{rest}}) = \sum_{i=1}^{k} \alpha_i \boldsymbol{\mu}_i = \sum_{i=1}^{k} \alpha_i \boldsymbol{\mu}_0 = \boldsymbol{\mu}_0 = \mathrm{E}(\mathbf{y}) \ . \tag{4.25}$$

Again we see that under the restriction $\sum_{i=1}^{k} \alpha_i = 1$ the same weights α_i occur in the representation of the combined forecast error in terms of the individual forecast errors:

$$\mathbf{e}_{\boldsymbol{\alpha},\text{rest}} = \mathbf{f}_{\boldsymbol{\alpha},\text{rest}} - \mathbf{y} = \sum_{i=1}^{k} \alpha_i \mathbf{f}_i - \left(\sum_{i=1}^{k} \alpha_i \right) \mathbf{y} = \sum_{i=1}^{k} \alpha_i (\mathbf{f}_i - \mathbf{y}) = \sum_{i=1}^{k} \alpha_i \mathbf{e}_i \ . \tag{4.26}$$

Using the restriction again we arrive at

$$\mathbf{e}_{\boldsymbol{\alpha},\text{rest}} = \left(1 - \sum_{i=2}^{k} \alpha_i \right) \mathbf{e}_1 + \sum_{i=2}^{k} \alpha_i \mathbf{e}_i = \mathbf{e}_1 + \sum_{i=2}^{k} \alpha_i (\mathbf{e}_i - \mathbf{e}_1) \ . \tag{4.27}$$

Since the individual forecasts are unbiased we see that $\mathbf{e}_1$ is an unbiased statistic for the non-stochastic $\mathbf{0} \in \mathbb{R}^l$ and that $\mathbf{e}_i - \mathbf{e}_1$ are unbiased for $\mathbf{0} \in \mathbb{R}^l$ for $i = 2, \dots, k$. In this situation we can apply the weak covariance adjustment technique from Lemma A.8 with $\mathbf{T} = \mathbf{e}_1$, $\boldsymbol{\theta} = \mathbf{0} \in \mathbb{R}^l$, $p = k-1$, $\mathbf{Z}_i = \mathbf{e}_{i+1} - \mathbf{e}_1$ and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)^{\mathrm{T}} = -(\alpha_2, \dots, \alpha_k)^{\mathrm{T}}$.
Note that $(\mathbf{Z}_1^{\mathrm{T}}, \dots, \mathbf{Z}_{k-1}^{\mathrm{T}})^{\mathrm{T}} = \mathbf{d}$ from Equation (3.4) such that

$$\begin{pmatrix} \mathrm{Cov}(\mathbf{Z}_1) & \mathrm{Cov}(\mathbf{Z}_1, \mathbf{Z}_2) & \dots & \mathrm{Cov}(\mathbf{Z}_1, \mathbf{Z}_{k-1}) \\ \mathrm{Cov}(\mathbf{Z}_2, \mathbf{Z}_1) & \mathrm{Cov}(\mathbf{Z}_2) & \dots & \mathrm{Cov}(\mathbf{Z}_2, \mathbf{Z}_{k-1}) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{Cov}(\mathbf{Z}_{k-1}, \mathbf{Z}_1) & \mathrm{Cov}(\mathbf{Z}_{k-1}, \mathbf{Z}_2) & \dots & \mathrm{Cov}(\mathbf{Z}_{k-1}) \end{pmatrix} = \mathrm{Cov}(\mathbf{d}) = \breve{\mathbf{W}}_{22} \tag{4.28}$$

from Equation (3.14) and

$$(\mathrm{Cov}(\mathbf{T}, \mathbf{Z}_1) | \dots | \mathrm{Cov}(\mathbf{T}, \mathbf{Z}_{k-1})) = \mathrm{Cov}(\mathbf{T}, \mathbf{d}) = \breve{\mathbf{W}}_{12} \tag{4.29}$$

from Equation (3.13). Remember that the $\breve{}$-accent denotes calculation under the unbiasedness assumption.
Taking traces of the respective $l \times l$-submatrices of $\breve{\mathbf{W}}_{12}$ and $\breve{\mathbf{W}}_{22}$ we obtain the vector

$$\breve{\mathbf{g}} = (\mathrm{tr}[\mathrm{Cov}(\mathbf{T}, \mathbf{Z}_1)], \dots, \mathrm{tr}[\mathrm{Cov}(\mathbf{T}, \mathbf{Z}_{k-1})])^{\mathrm{T}} \tag{4.30}$$

and the matrix

$$\breve{\mathbf{G}} = (\mathrm{tr}[\mathrm{Cov}(\mathbf{Z}_i, \mathbf{Z}_s)])_{i,s=1,\dots,k-1} \ . \tag{4.31}$$

From Lemma A.8 and the restriction $\mathbf{1}_k^{\mathrm{T}}\boldsymbol{\alpha} = 1$ we get the optimal choice for $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)^{\mathrm{T}}$ as

$$\begin{pmatrix} \alpha_{2,\mathrm{opt}} \\ \vdots \\ \alpha_{k,\mathrm{opt}} \end{pmatrix} = -\boldsymbol{\gamma}_{\mathrm{opt}} = -\breve{\mathbf{G}}^{-1}\breve{\mathbf{g}} \quad \text{and} \quad \alpha_{1,\mathrm{opt}} = 1 - \sum_{i=2}^{k} \alpha_{i,\mathrm{opt}} \; . \tag{4.32}$$

In deriving the corresponding optimal value of the SMSPE-function again we have to distinguish between two situations: If the individual forecasts $\mathbf{f}_i$ and hence also $\mathbf{f}_{\boldsymbol{\alpha},\mathrm{rest}}$ are actually unbiased, we have $\mathrm{E}(\mathbf{e}_{\boldsymbol{\alpha},\mathrm{rest}}) = \mathbf{0}$ and, thus,

$$\mathrm{tr}(\mathrm{Cov}(\mathbf{e}_{\boldsymbol{\alpha},\mathrm{rest}})) = \mathrm{E}[(\mathbf{e}_{\boldsymbol{\alpha},\mathrm{rest}})^{\mathrm{T}}(\mathbf{e}_{\boldsymbol{\alpha},\mathrm{rest}})] = \mathrm{SMSPE}(\mathbf{f}_{\boldsymbol{\alpha},\mathrm{rest}}, \mathbf{y}) \; . \tag{4.33}$$

Consequently, from Lemma A.8 we have the optimal value of the SMSPE-function

$$\mathrm{SMSPE}(\mathbf{f}_{\boldsymbol{\alpha}_{\mathrm{opt}},\mathrm{rest}}, \mathbf{y}) = \mathrm{tr}(\breve{\mathbf{W}}_{11}) - \breve{\mathbf{g}}^{\mathrm{T}}\breve{\mathbf{G}}^{-1}\breve{\mathbf{g}} \; , \tag{4.34}$$

where $\breve{\mathbf{W}}_{11} = \mathrm{Cov}(\mathbf{e}_1)$. If, however, we cannot assume unbiasedness we have to calculate the SMSPE-value by inserting $\mathbf{B} = \boldsymbol{\alpha}_{\mathrm{opt}}^{\mathrm{T}} \otimes \mathbf{I}_l$ and $\mathbf{c} = \mathbf{0}$ into the general SMSPE-function from Equation (2.5).

As the last possibility of the weak approach we consider the combined forecast

$$\mathbf{f}_{\boldsymbol{\alpha},\mathbf{c},\mathrm{rest}} = \sum_{i=1}^{k} \alpha_i \mathbf{f}_i + \mathbf{c} \quad \text{with} \quad \sum_{i=1}^{k} \alpha_i = 1 \; , \tag{4.35}$$

where unbiasedness of the individual forecasts is no longer assumed.
Following the same reasoning as in Section 3 we can conclude that the optimal parameter values in this case are given by

$$\mathbf{c}_{\mathrm{opt}} = \boldsymbol{\mu}_0 - \sum_{i=1}^{k} \alpha_{i,\mathrm{opt}} \boldsymbol{\mu}_i \; , \tag{4.36}$$

$$\begin{pmatrix} \alpha_{2,\mathrm{opt}} \\ \vdots \\ \alpha_{k,\mathrm{opt}} \end{pmatrix} = -\mathbf{G}^{-1}\mathbf{g} \quad \text{and} \quad \alpha_{1,\mathrm{opt}} = 1 - \sum_{i=2}^{k} \alpha_{i,\mathrm{opt}} \; , \tag{4.37}$$

whereas the corresponding optimal value of the SMSPE-function is

$$\mathrm{SMSPE}(\mathbf{f}_{\boldsymbol{\alpha}_{\mathrm{opt}},\mathbf{c}_{\mathrm{opt}},\mathrm{rest}}, \mathbf{y}) = \mathrm{tr}(\mathbf{W}_{11}) - \mathbf{g}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{g} \; . \tag{4.38}$$

Here all moments have to be calculated without the unbiasedness assumption. $\mathbf{g}$ and $\mathbf{G}$ are calculated from $\mathbf{W}_{12}$ and $\mathbf{W}_{22}$ in exactly the same way as $\breve{\mathbf{g}}$ and $\breve{\mathbf{G}}$ are calculated from $\breve{\mathbf{W}}_{12}$ and $\breve{\mathbf{W}}_{22}$.

It should be pointed out, that the weak combination approach needs even less parameters than would be involved if the forecasts for each target variable y_j were combined within the univariate setting (which is the medium approach as outlined above). Consequently, it may be practical in empirical applications where the number of data available for parameter estimation is not large.
On the other hand all components of the forecasts are treated alike by the weak approach: The same coefficient α_i is assigned to each component of the forecast $\mathbf{f}_i$. Thus α_i constitutes a compromise between the choices which would have been made for each component separately in the medium approach. Only the constant vector $\mathbf{c}$ allows for an individual correction in each of the l components. In how far such a procedure is reasonable depends on the variables under consideration.

4.3 Arithmetic mean and general remarks

Finally, we will also include the arithmetic mean of the individual forecasts in our considerations: It is a very simple linear combination and has proven (at least in the univariate case) to be very powerful in empirical investigations (cf. Kang, 1986).

$$\begin{aligned}\mathbf{f}_{\text{am}} = \frac{1}{k}\sum_{i=1}^{k}\mathbf{f}_i &= \mathbf{Bf} \quad \text{with} \quad \mathbf{B} = \frac{1}{k}(\mathbf{I}_l|\dots|\mathbf{I}_l) \in \mathbb{R}^{l\times kl} \\ &= \mathbf{f}_{\boldsymbol{\alpha}} \quad \text{with} \quad \boldsymbol{\alpha} = \frac{1}{k}\mathbf{1}_k \;. \end{aligned} \tag{4.39}$$

Its expectation is

$$\mathrm{E}(\mathbf{f}_{\text{am}}) = \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\mu}_i \tag{4.40}$$

and thus the unweighted average is not unbiased in general. If, however, each individual forecast is unbiased, then also $\mathbf{f}_{\text{am}}$ is. The corresponding SMSPE-value can be obtained by inserting $\boldsymbol{\alpha} = \frac{1}{k}\mathbf{1}_k$ and $\mathbf{c} = \mathbf{0}$ into Equation (4.9):

$$\mathrm{SMSPE}(\mathbf{f}_{\text{am}}, \mathbf{y}) = \frac{1}{k^2}\mathbf{1}_k^{\mathrm{T}}\mathbf{H}\mathbf{1}_k - \frac{2}{k}\mathbf{1}_k^{\mathrm{T}}\mathbf{h} + \mathrm{tr}(\boldsymbol{\Sigma}_{00}) + \boldsymbol{\mu}_0^{\mathrm{T}}\boldsymbol{\mu}_0 \;. \tag{4.41}$$

Taking a closer look at the optimal combination parameters in the strong, medium and weak multivariate linear approaches we detect that all of them utilize moments

of the joint distribution of $\mathbf{y}$ and $\mathbf{f}$ up to order two. The *level* of knowledge required, however, is different: The strong approaches work with the full matrices $\mathbf{\Sigma}_{is}$ $(i, s = 0, \ldots, k)$, whereas the medium and weak approaches only use the diagonal elements of the $\mathbf{\Sigma}_{is}$. While the medium approaches use these diagonal elements individually, the weak approaches only work with their sums, i.e. with the traces of the $\mathbf{\Sigma}_{is}$.
It is important to note that we may as well regard the problem of finding the optimal combination parameters in the multivariate linear approaches as linear regression problems just like Granger and Ramanathan (1984) showed for the univariate case. This point of view is very practical for empirical applications since standard statistics software can be used. For details on the regression interpretation we refer to Troschke (2002).
In Section 6 we will report about first investigations on the quality of the multivariate linear approaches in the simple case of $k = 2$ forecasts for $l = 2$ variables. But first we will turn to the special case $k = 1$.

5 The special case $k = 1$: Adjustment of forecasts

There is no reason why the special case $k = 1$ should be ruled out in the above considerations. Of course, this "combination of one multivariate forecast" should rather be addressed as *adjustment of single multivariate forecasts*. Exploiting the moment structure of the joint distribution of the target vector variable $\mathbf{y}$ and a single forecast $\mathbf{f}_i$ the performance of $\mathbf{f}_i$ can be improved with respect to the scalar mean square prediction error by this kind of adjustment. Granger (1989, p. 169) points out the usefulness of such adjustments in the univariate case.
The SMSPE of the forecast $\mathbf{f}_i$ is given by

$$\begin{aligned}
\mathrm{SMSPE}(\mathbf{f}_i, \mathbf{y}) &= \mathrm{E}[(\mathbf{f}_i - \mathbf{y})^{\mathrm{T}}(\mathbf{f}_i - \mathbf{y})] \\
&= \mathrm{tr}[\mathrm{Cov}(\mathbf{f}_i - \mathbf{y})] + [\mathrm{E}(\mathbf{f}_i - \mathbf{y})]^{\mathrm{T}}[\mathrm{E}(\mathbf{f}_i - \mathbf{y})] \\
&= \mathrm{tr}(\mathbf{\Sigma}_{ii}) + \boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\mu}_i + \mathrm{tr}(\mathbf{\Sigma}_{00}) + \boldsymbol{\mu}_0^{\mathrm{T}}\boldsymbol{\mu}_0 - 2[\mathrm{tr}(\mathbf{\Sigma}_{i0}) + \boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\mu}_0]\ . \quad (5.1)
\end{aligned}$$

All of the multivariate linear combination approaches described above may be employed in this case resulting in the corresponding adjustments. Some of the adjustments, however, are identical to others: While $\mathbf{f}_{\mathbf{B},\mathbf{c},\mathrm{rest}}$, $\mathbf{f}_{\mathbf{D},\mathbf{c},\mathrm{rest}}$ and $\mathbf{f}_{\boldsymbol{\alpha},\mathbf{c},\mathrm{rest}}$ coincide for $k = 1$, the adjustment counterparts of $\mathbf{f}_{\mathbf{B},\mathrm{rest}}$, $\mathbf{f}_{\mathbf{D},\mathrm{rest}}$ and $\mathbf{f}_{\boldsymbol{\alpha},\mathrm{rest}}$ coincide with the individual forecast. Consequently, there are only seven remaining different adjustments.
The optimal adjustment parameters and their corresponding optimal SMSPE-values may be obtained from the respective formulae in the preceding sections by setting

$\mathbf{f} = \mathbf{f}_i$, $\boldsymbol{\mu}_\mathbf{f} = \boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_{\mathbf{ff}} = \boldsymbol{\Sigma}_{ii}$, $\boldsymbol{\Sigma}_{0\mathbf{f}} = \boldsymbol{\Sigma}_{0i}$, $\mathbf{B} = \mathbf{B}_i \in \mathbb{R}^{l\times l}$, $\mathbf{D} = \mathbf{D}_i \in \mathbb{R}^{l\times l}$ and $\boldsymbol{\alpha} = \alpha_i \in \mathbb{R}$. We will illustrate this by two examples:
The *strong multivariate unrestricted linear adjustment of* $\mathbf{f}_i$ *with constant term* is

$$(\mathbf{f}_i)_{\mathbf{B},\mathbf{c}} = \mathbf{B}\mathbf{f}_i + \mathbf{c} \ , \tag{5.2}$$

where $\mathbf{B} \in \mathbb{R}^{l\times l}$ and $\mathbf{c} \in \mathbb{R}^l$. The optimal choices for the parameters are obtained as special cases of Equations (2.9) and (2.10), namely

$$\mathbf{B}_{\text{opt}} = \boldsymbol{\Sigma}_{0i}\boldsymbol{\Sigma}_{ii}^{-1} \quad \text{and} \quad \mathbf{c}_{\text{opt}} = \boldsymbol{\mu}_0 - \boldsymbol{\Sigma}_{0i}\boldsymbol{\Sigma}_{ii}^{-1}\boldsymbol{\mu}_i \tag{5.3}$$

with corresponding optimal SMSPE-value from Equation (2.12)

$$\text{SMSPE}((\mathbf{f}_i)_{\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}}}, \mathbf{y}) = \text{tr}(\boldsymbol{\Sigma}_{00}) - \text{tr}(\boldsymbol{\Sigma}_{0i}\boldsymbol{\Sigma}_{ii}^{-1}\boldsymbol{\Sigma}_{i0}) \ . \tag{5.4}$$

The *strong (and medium and weak) multivariate linear adjustment of* $\mathbf{f}_i$ *with constant term and with the restriction of the weights summing up to the identity matrix* is

$$(\mathbf{f}_i)_{\mathbf{I}_l,\mathbf{c}} = \mathbf{f}_i + \mathbf{c} \ . \tag{5.5}$$

According to Equation (3.21) the optimal choice for $\mathbf{c} \in \mathbb{R}^l$ is given by

$$\mathbf{c}_{\text{opt}} = \boldsymbol{\mu}_0 - \boldsymbol{\mu}_i \ , \tag{5.6}$$

thus resulting in the bias corrected forecast. The corresponding optimal SMSPE-value is

$$\text{SMSPE}((\mathbf{f}_i)_{\mathbf{I}_l,\mathbf{c}_{\text{opt}}}, \mathbf{y}) = \text{tr}(\boldsymbol{\Sigma}_{ii}) + \text{tr}(\boldsymbol{\Sigma}_{00}) - 2\,\text{tr}(\boldsymbol{\Sigma}_{i0}) \ . \tag{5.7}$$

As a consequence of the previous results each of the adjusted forecasts employing a constant term $\mathbf{c}$ is unbiased.
In the following Section 6 we will carry out a first analysis of the potential inherent in the new methods. The seven different ways of adjusting a single forecast will be included.

6 First comparisons

The purpose of this section is to present a first example of the potential inherent in the multivariate adjustments and combinations. We will consider the simple case of $k = 2$ forecasts $\mathbf{f}_1$ and $\mathbf{f}_2$ for an $l = 2$ dimensional target variable $\mathbf{y}$. Consequently, the number of unknown combination parameters ranges from 1 for the weak $\mathbf{f}_{\boldsymbol{\alpha},\text{rest}}$ combination to 10 for the strong $\mathbf{f}_{\mathbf{B},\mathbf{c}}$ combination.

The example consists of a comparison of the optimal SMSPE-values which result from a given set of moments $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. In order to base the example on realistic grounds these moments were obtained as the sample moments from an empirical forecast data set. Of course such an example can provide only limited insight in the performance of the various methods. Therefore, a more detailed analysis has to follow and will be presented in a future paper.

We assume $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to be given as follows:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_0 \\ \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} = \left(\begin{array}{c} 2.328571 \\ 1.961905 \\ \hline 1.904762 \\ 1.857143 \\ \hline 2.047619 \\ 1.928571 \end{array}\right) \tag{6.1}$$

and

$$\begin{aligned} \boldsymbol{\Sigma} &= \begin{pmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} & \boldsymbol{\Sigma}_{02} \\ \boldsymbol{\Sigma}_{10} & \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{20} & \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \\ &= \left(\begin{array}{cc|cc|cc} 3.018231 & 2.335850 & 2.412245 & 1.777891 & 1.899830 & 1.621088 \\ 2.335850 & 2.938549 & 2.265420 & 2.246939 & 1.812528 & 1.985374 \\ \hline 2.412245 & 2.265420 & 3.229025 & 2.272109 & 2.510488 & 2.088435 \\ 1.777891 & 2.246939 & 2.272109 & 2.622449 & 1.649660 & 2.335034 \\ \hline 1.899830 & 1.812528 & 2.510488 & 1.649660 & 2.134637 & 1.562925 \\ 1.621088 & 1.985374 & 2.088435 & 2.335034 & 1.562925 & 2.221088 \end{array}\right) \end{aligned} \tag{6.2}$$

From these moments we may now determine the optimal adjustment or combination parameters belonging to the different methods. In order to obtain the corresponding optimal SMSPE-values we can use the formulae derived in Sections 2 - 5.

Following the formulae from Section 2 we obtain, for example, the optimal parameters for the strong multivariate linear unrestricted combination approach with constant term $\mathbf{f}_{\mathbf{B},\mathbf{c}}$:

$$\begin{aligned} \mathbf{B}_{\text{opt}} &= \begin{pmatrix} 0.505969 & 0.199559 & 0.223352 & -0.112853 \\ -0.448593 & 1.124554 & 0.845578 & -0.461582 \end{pmatrix} \quad \text{and} \\ \mathbf{c}_{\text{opt}} &= \begin{pmatrix} 0.754516 \\ -0.113317 \end{pmatrix} . \end{aligned} \tag{6.3}$$

The corresponding optimal value of the SMSPE-function is

$$\text{SMSPE}(\mathbf{f}_{\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}}}, \mathbf{y}) = 0.800234 \, . \tag{6.4}$$

Since the arithmetic mean of the individual forecasts is a very simple and empirically very successful combining strategy, we decided to compare the potential of the new techniques to that of the arithmetic mean. Consequently, all SMSPE-values in Table 1 are presented relative to the SMSPE-value of the arithmetic mean, which is 2.515893 in the present situation. All decimals have been deleted after the fourth decimal such that methods outperforming the arithmetic mean can be identified immediately.

Since the moments $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are assumed to be known, the calculations are carried out on a theoretical basis and, hence, the SMSPE-values reflect the theoretical ranking of the various methods: the adjustments are not worse than the single forecasts; the more elaborate combinations are not worse than the arithmetic mean and they are not worse than the individual forecasts; the combination $f_{\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}}}$ is not worse than $f_{\mathbf{B}_{\text{opt}}}$ which in turn is not worse than $f_{\mathbf{B}_{\text{opt}},\text{rest}}$; likewise $f_{\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}}}$ is not worse than $f_{\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}},\text{rest}}$ which in turn is not worse than $f_{\mathbf{B}_{\text{opt}},\text{rest}}$; the strong version of a linear combination is not worse than the corresponding medium version which in turn is not worse than the corresponding weak version.

It is interesting to analyze what margin of improvement is expected in the situation under consideration: The expected scalar squared error loss of the best combination $f_{\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}}}$ is 20% less than that of the arithmetic mean. The corresponding medium and weak combinations are expected to be only 11% better than the arithmetic mean. The medium approach is equivalent to the classical univariate treatment of each component. Consequently, we can infer that the strong multivariate treatment (incorporating the interactions between the components) has the potential to improve upon the classical univariate treatment by 10% in the current example. The best adjustments $f_{i,\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}}}$ are expected to be about 20% or 13% better than their corresponding individual forecasts. It is also interesting to note that in general the strong combinations and adjustments are far better than their medium counterparts whereas only small differences can be observed between the medium and the weak counterparts. One should, however, take into consideration that the medium and weak combinations and adjustments depend on far fewer parameters than their strong counterparts. This may result in an advantage, when the combination or adjustment parameters have to be estimated.

The results also reveal that the expected advantage of including a constant term and / or dropping the restriction can be substantial.

Forecast $f_\cdot$		SMSPE$(f_\cdot, y)$
First forecast	$\mathbf{f}_1$	1.0654
Adjustments:	$f_{1,\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}}}$	0.8516
	$f_{1,\mathbf{B}_{\text{opt}}}$	0.9909
	$f_{1,\mathbf{I},\mathbf{c}_{\text{opt}}}$	0.9896
	$f_{1,\mathbf{D}_{\text{opt}},\mathbf{c}_{\text{opt}}}$	0.8861
	$f_{1,\mathbf{D}_{\text{opt}}}$	1.0632
	$f_{1,\alpha_{\text{opt}},\mathbf{c}_{\text{opt}}}$	0.8931
	$f_{1,\alpha_{\text{opt}}}$	1.0643
Second forecast	$\mathbf{f}_2$	1.0422
Adjustments:	$f_{2,\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}}}$	0.9043
	$f_{2,\mathbf{B}_{\text{opt}}}$	0.9263
	$f_{2,\mathbf{I},\mathbf{c}_{\text{opt}}}$	1.0104
	$f_{2,\mathbf{D}_{\text{opt}},\mathbf{c}_{\text{opt}}}$	0.9902
	$f_{2,\mathbf{D}_{\text{opt}}}$	1.0329
	$f_{2,\alpha_{\text{opt}},\mathbf{c}_{\text{opt}}}$	0.9902
	$f_{2,\alpha_{\text{opt}}}$	1.0413
Strong combinations:	$f_{\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}}}$	0.8002
	$f_{\mathbf{B}_{\text{opt}}}$	0.8483
	$f_{\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}},\text{rest}}$	0.9030
	$f_{\mathbf{B}_{\text{opt}},\text{rest}}$	0.9478
Medium combinations:	$f_{\mathbf{D}_{\text{opt}},\mathbf{c}_{\text{opt}}}$	0.8842
	$f_{\mathbf{D}_{\text{opt}}}$	0.9771
	$f_{\mathbf{D}_{\text{opt}},\mathbf{c}_{\text{opt}},\text{rest}}$	0.9388
	$f_{\mathbf{D}_{\text{opt}},\text{rest}}$	0.9851
Weak combinations:	$f_{\boldsymbol{\alpha}_{\text{opt}},\mathbf{c}_{\text{opt}}}$	0.8922
	$f_{\boldsymbol{\alpha}_{\text{opt}}}$	0.9987
	$f_{\boldsymbol{\alpha}_{\text{opt}},\mathbf{c}_{\text{opt}},\text{rest}}$	0.9482
	$f_{\boldsymbol{\alpha}_{\text{opt}},\text{rest}}$	0.9993

Table 1: SMSPE-values of adjusted and combined forecasts for certain known moments $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (all values relative to the SMSPE of the arithmetic mean)

It can be seen that there is some potential in the multivariate linear approaches to outperform the arithmetic mean. How well this potential is exploited will depend on how good the current sample reflects the future relationship between target variable $\mathbf{y}$ and forecasts $\mathbf{f}_i$. Clearly, the more suitable data are available for estimating that relationship, the better. Also the data should not be subject to extreme structural changes during the considered period. Consequently, the multivariate linear approaches should be more valuable for monthly, weekly or even daily data (e.g. from the stock market) than they are for yearly data.

7 Translations and Scale Transformations

It is an important thing to know in which way the linear adjustments and combinations of multivariate forecasts are affected by transformations of origin and scale, i.e. in how far the results depend on the chosen coordinate system. We investigate if the adjusted and combined forecasts are transformed in the same way as the individual forecasts.

Let us first consider translations of the data. By this we mean that we add a constant vector $\boldsymbol{\tau} \in \mathbb{R}^l$ to the target variable $\mathbf{y}$ as well as to each single forecast $\mathbf{f}_i$, i.e. after the translation we obtain the new variables

$$\begin{pmatrix} \tilde{\mathbf{y}} \\ \tilde{\mathbf{f}} \end{pmatrix} = \begin{pmatrix} \mathbf{y} \\ \mathbf{f} \end{pmatrix} + \mathbf{1}_{k+1} \otimes \boldsymbol{\tau} \ . \tag{7.1}$$

The expectation vector $\tilde{\boldsymbol{\mu}}$ and the centered second order moment matrix $\tilde{\boldsymbol{\Sigma}}$ of the transformed variables $(\tilde{\mathbf{y}}^{\mathrm{T}}, \tilde{\mathbf{f}}^{\mathrm{T}})^{\mathrm{T}}$ relate to the corresponding quantities $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of the original variables $(\mathbf{y}^{\mathrm{T}}, \mathbf{f}^{\mathrm{T}})^{\mathrm{T}}$ as follows:

$$\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu} + \mathbf{1}_{k+1} \otimes \boldsymbol{\tau} \quad \text{and} \quad \tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} \ . \tag{7.2}$$

The second order moment matrix $\breve{\boldsymbol{\Sigma}}$ which is calculated differently because of the assumption of unbiasedness of each single forecast is not affected by such a translation either. Consequently, also the matrices $\mathbf{V}$, $\breve{\mathbf{V}}$, $\mathbf{W}$ and $\breve{\mathbf{W}}$ are not affected. The matrices and vectors $\mathbf{H}_1$, $\mathbf{h}_1$, $\mathbf{G}$, $\mathbf{g}$, $\breve{\mathbf{G}}$ and $\breve{\mathbf{g}}$ connected with the weak approaches are not affected as well, whereas $\mathbf{H}$ and $\mathbf{h}$ are changed.

Consulting the equations determining the optimal parameters from the respective sections above we can derive that all combinations except three behave reasonably under the translation, i.e. they are translated by the same vector $\boldsymbol{\tau}$ as the individual forecasts. The exceptions are $\mathbf{f}_{\mathbf{B}_{\mathrm{opt}}}$, $\mathbf{f}_{\mathbf{D}_{\mathrm{opt}}}$ and $\mathbf{f}_{\boldsymbol{\alpha}_{\mathrm{opt}}}$. The corresponding adjustments show the same behaviour. Of course, the arithmetic mean is translated by $\boldsymbol{\tau}$ as well.

Let us now turn to scale transformations of the data. By this we mean that target variable $\mathbf{y}$ as well as each single forecast $\mathbf{f}_i$ are multiplied by the same constant matrix $\mathbf{\Lambda} \in \mathbb{R}^{l \times l}$, i.e. after the translation we obtain the new variables

$$\begin{pmatrix} \tilde{\mathbf{y}} \\ \tilde{\mathbf{f}} \end{pmatrix} = \begin{pmatrix} \mathbf{\Lambda y} \\ \mathbf{\Lambda f}_1 \\ \vdots \\ \mathbf{\Lambda f}_k \end{pmatrix} = \begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{\Lambda} \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_k \end{pmatrix} =: \mathbf{\Gamma} \begin{pmatrix} \mathbf{y} \\ \mathbf{f} \end{pmatrix} . \tag{7.3}$$

The moments of the transformed variables $(\tilde{\mathbf{y}}^{\mathrm{T}}, \tilde{\mathbf{f}}^{\mathrm{T}})^{\mathrm{T}}$ relate to the corresponding quantities of the original variables $(\mathbf{y}^{\mathrm{T}}, \mathbf{f}^{\mathrm{T}})^{\mathrm{T}}$ as follows:

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_i &= \mathbf{\Lambda} \boldsymbol{\mu}_i \quad \text{for } i = 0, \dots, k \qquad \text{and} \\ \tilde{\mathbf{\Sigma}}_{is} &= \mathbf{\Lambda \Sigma}_{is} \mathbf{\Lambda}^{\mathrm{T}} \quad \text{for } i, s = 0, \dots, k , \end{aligned} \tag{7.4}$$

i.e. $\tilde{\boldsymbol{\mu}} = \mathbf{\Gamma} \boldsymbol{\mu}$ and $\tilde{\mathbf{\Sigma}} = \mathbf{\Gamma \Sigma \Gamma}^{\mathrm{T}}$. The $l \times l$ submatrices of $\breve{\mathbf{\Sigma}}$, $\mathbf{V}$, $\breve{\mathbf{V}}$, $\mathbf{W}$ and $\breve{\mathbf{W}}$ are transformed in the same way.

Analyzing the respective formulae we can conclude that all strong combinations and adjustments as well as the arithmetic mean show a reasonable behaviour under such a transformation: They are multiplied by $\mathbf{\Lambda}$ from the left as well. The medium combinations and adjustments only behave reasonably if we restrict $\mathbf{\Lambda}$ to be a diagonal matrix, whereas the weak combinations and adjustments perform well only if we further restrict $\mathbf{\Lambda} = \lambda \mathbf{I}_l$, $\lambda \in \mathbb{R}$. This exactly reflects the spirit of the three approaches, since a diagonal $\mathbf{\Lambda}$ means treatment of each component on its own, whereas a scalar matrix $\mathbf{\Lambda}$ means treatment of all components as one. In the latter case we may conclude that $\tilde{\mathbf{H}}_1 = \lambda^2 \mathbf{H}_1$ and the vectors and matrices $\mathbf{h}_1$, $\mathbf{H}$, $\mathbf{h}$, $\mathbf{G}$, $\mathbf{g}$, $\breve{\mathbf{G}}$ and $\breve{\mathbf{g}}$ are transformed in the same way.

Combining the results on translations and scale transformations above we may conclude that only the combinations $\mathbf{f}_{\mathbf{B}_{\mathrm{opt}}, \mathbf{c}_{\mathrm{opt}}}$, $\mathbf{f}_{\mathbf{B}_{\mathrm{opt}}, \mathrm{rest}}$, $\mathbf{f}_{\mathbf{B}_{\mathrm{opt}}, \mathbf{c}_{\mathrm{opt}}, \mathrm{rest}}$ and $\mathbf{f}_{am}$ and their corresponding adjustments remain reliable under linear transformations of the data. Thus, these combinations should be preferred if they show a good behaviour otherwise.

Besides their unreasonable behaviour under translations the combinations $\mathbf{f}_{\mathbf{B}_{\mathrm{opt}}}$, $\mathbf{f}_{\mathbf{D}_{\mathrm{opt}}}$ and $\mathbf{f}_{\boldsymbol{\alpha}_{\mathrm{opt}}}$ are not necessarily unbiased even if the individual forecasts are unbiased. Consequently, use of these techniques cannot be recommended.

8 Conclusions

In this report we have introduced linear approaches for the combination of *multivariate* forecasts. Three main variants of this approach have been considered. The strong variant depends on the largest number of unknown combination parameters followed by the medium and then the weak variant. Within each main variant we have considered four further variants. They can be characterized by two facts well-known from the univariate linear combination of forecasts: "Are the weight matrices restricted to sum up to the identity matrix?" and "Is a constant term included?" For each case we have derived the respective optimal combination parameters as well as the resulting optimal value of the scalar mean square prediction error.

Each of the multivariate linear approaches requires knowledge about the moments up to order two of the joint distribution of $\mathbf{y}$ and $\mathbf{f}$. Again, the strong variant requires more knowledge than the medium variant, and the medium variant requires more detailed knowledge than the weak variant. Since the medium approach is equivalent to the univariate treatment of each component, the classical univariate linear approaches have been included as competitors to the new approaches. We have also considered the special case of $k = 1$ forecast which means adjustment of an individual forecast.

Due to the smaller number of parameters involved the weak combination and adjustment approaches may be apt if only a small amount of data is available for parameter estimation. Use of the weak approaches, however, amounts to the fact that all components are treated as one. Thus, they should only be applied if this seems at least roughly justified.

We have seen that the strong multivariate linear combinations $f_{\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}}}$, $f_{\mathbf{B}_{\text{opt}},\mathbf{c}_{\text{opt}},\text{rest}}$ and $f_{\mathbf{B}_{\text{opt}},\text{rest}}$ as well as the arithmetic mean show a reasonable behaviour when the coordinate system is changed in which the target variable and the forecasts are measured. From this point of view use of the strong, medium and weak linear unrestricted combinations of forecasts without constant term $f_{\mathbf{B}_{\text{opt}}}$, $f_{\mathbf{D}_{\text{opt}}}$ and $f_{\boldsymbol{\alpha}_{\text{opt}}}$ are the least advisable.

Granger and Ramanathan (1984) have shown that in the univariate case the linear combination of forecasts can be seen as a linear regression problem. The same is true for the linear combination of multivariate forecasts introduced in this paper. The appropriate regression models will be derived in a further paper (Troschke, 2001).

An analysis of the potential of the various methods with respect to their optimal expected scalar squared prediction error loss (SMSPE) was carried out. We have considered a small example based on given realistic moments of the joint distribution of $\mathbf{y}$ and $\mathbf{f}$. In this example the possible improvement with respect to the classical univariate approach was 10% whereas the possible improvement with respect to the arithmetic mean was 20%. If the moments have to be estimated some portion of that advantage will be lost. How much will be lost with a certain sample size is one question which has to be answered by a much more detailed analysis of the possible benefits of the multivariate linear approaches. Another point is to find out whether it is worthwhile to consider more than $k = 2$ forecasts or more than $l = 2$ variables at a time.

Appendix

A Preliminary results

This section collects some results which are needed for our considerations. Most of them are well-known from the literature.
The first lemma provides the inverse of a regular matrix modified by a matrix of rank one:

Lemma A.1 *(*Rao *and* Bhimasankaram*, 1992, p. 145) Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *be non-singular and let* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$*. Then*

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^{\mathrm{T}})^{-1} = \mathbf{A}^{-1} - \frac{1}{1 + \mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{u}}\, \mathbf{A}^{-1}\mathbf{u}\mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1} \,.$$

The following lemma gives explicit representations of some matrix or vector expressions in terms of the elements involved.

Lemma A.2 *Let* $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{m\times n}$*,* $\mathbf{x} = (x_i) \in \mathbb{R}^m$ *and* $\mathbf{y} = (y_j) \in \mathbb{R}^n$*. Then*

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{y} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_i y_j \,.$$

In the special case where $m = n$ *and* $\mathbf{A} = \mathbf{I}_n$ *we obtain*

$$\mathbf{x}^{\mathrm{T}}\mathbf{y} = \sum_{i=1}^{n} x_i y_i \,.$$

In order to determine the optimal combination parameters within our various approaches we have to apply differential calculus. Since some of the parameters are vectors or even matrices the concept of matrix differential calculus (Magnus and Neudecker, 1999) proves most helpful.

Definition A.3 *Let* $f(\mathbf{X})$ *be a real-valued function of a matrix* $\mathbf{X} = (x_{ij}) \in \mathbb{R}^{n\times q}$*. Then* f *is called differentiable with respect to* $\mathbf{X}$ *if and only if it is differentiable with respect to each of the elements* x_{ij}*. The derivative of* f *with respect to* $\mathbf{X}$

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} := \begin{pmatrix} \partial f(\mathbf{X})/\partial x_{11} & \dots & \partial f(\mathbf{X})/\partial x_{1q} \\ \vdots & & \vdots \\ \partial f(\mathbf{X})/\partial x_{n1} & \dots & \partial f(\mathbf{X})/\partial x_{nq} \end{pmatrix}$$

is a matrix with the same dimensions as $\mathbf{X}$*.*

Lemma A.4 *(*MAGNUS *and* NEUDECKER*, 1999, pp. 119f, 95, 65) Let $f : S \to \mathbb{R}$ be a real-valued function defined on a set $S \subset \mathbb{R}^{n \times q}$. Further let $\mathbf{X}_0$ be an interior point of S and let f be differentiable at $\mathbf{X}_0$. Then a necessary condition for f to have a local minimum or a local maximum at $\mathbf{X}_0$ is*

$$\frac{\partial f(\mathbf{X}_0)}{\partial \mathbf{X}} = \mathbf{0} \, ,$$

where the derivative of f with respect to $\mathbf{X}$ is given in Definition A.3 above.
$\mathbf{X}_0$ is an interior point of S if there exists a real constant r such that the ball

$$B(\mathbf{X}_0; r) = \{\mathbf{X} | \mathbf{X} \in \mathbb{R}^{n \times q}, ||\mathbf{X} - \mathbf{X}_0|| < r\}$$

is contained in S completely. Here $||\mathbf{X}|| = (\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{X}))^{1/2}$ denotes the FROBENIUS-*norm of $\mathbf{X}$.*

The next two lemmas give the derivatives for special real-valued vector and matrix functions.

Lemma A.5 *(*MAGNUS *and* NEUDECKER*, 1999, p. 177) Let $\mathbf{a}, \mathbf{x} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{n \times n}$. Then*

$$\frac{\partial \mathbf{a}^{\mathrm{T}}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \, ,$$

$$\frac{\partial \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^{\mathrm{T}})\mathbf{x} \, .$$

Lemma A.6 *(*MAGNUS *and* NEUDECKER*, 1999, p. 178) Let $\mathbf{A}, \mathbf{B}, \mathbf{X}$ be real matrices of appropriate dimensions. Then*

$$\frac{\partial \, \mathrm{tr}(\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^{\mathrm{T}} \, ,$$

$$\frac{\partial \, \mathrm{tr}(\mathbf{X}\mathbf{A}\mathbf{X}^{\mathrm{T}}\mathbf{B})}{\partial \mathbf{X}} = \mathbf{B}^{\mathrm{T}}\mathbf{X}\mathbf{A}^{\mathrm{T}} + \mathbf{B}\mathbf{X}\mathbf{A} \, ,$$

$$\frac{\partial \, \mathrm{tr}(\mathbf{X}\mathbf{A}\mathbf{X}\mathbf{B})}{\partial \mathbf{X}} = \mathbf{B}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} + \mathbf{A}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}} \, .$$

The following two lemmas deal with strong and weak covariance adjustment.

Lemma A.7 (Strong covariance adjustment technique, RAO 1966, 1967, BAKSALARY and KALA 1982)
Let $\mathbf{T}$ *and* $\mathbf{Z}$ *be two statistics with* $\mathrm{E}(\mathbf{T}) = \boldsymbol{\theta} \in \mathbb{R}^l$, $\mathrm{E}(\mathbf{Z}) = \mathbf{0} \in \mathbb{R}^m$ *and*

$$\mathrm{Cov}\begin{pmatrix} \mathbf{T} \\ \mathbf{Z} \end{pmatrix} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{12}^{\mathrm{T}} & \mathbf{W}_{22} \end{pmatrix} = \mathbf{W}$$

where $\mathbf{W}$ *is a known positive definite matrix.*
Then

$$\mathbf{T}_0 = \mathbf{T} - \mathbf{W}_{12}\mathbf{W}_{22}^{-1}\mathbf{Z}$$

is unbiased for $\boldsymbol{\theta}$. *The dispersion matrix of* $\mathbf{T}_0$

$$\mathrm{Cov}(\mathbf{T}_0) = \mathbf{W}_{11} - \mathbf{W}_{12}\mathbf{W}_{22}^{-1}\mathbf{W}_{12}^{\mathrm{T}}$$

is minimal in the sense of the LÖWNER *ordering among all unbiased estimators for* $\boldsymbol{\theta}$, *which are linear combinations* $\mathbf{L}_1\mathbf{T} + \mathbf{L}_2\mathbf{Z}$ *with matrices* $\mathbf{L}_1$ *and* $\mathbf{L}_2$ *of appropriate dimensions.*

Lemma A.8 (Weak covariance adjustment technique, IHORST, 1993, TRENKLER and IHORST, 1995)
Let $\mathbf{T}$ *and* $\mathbf{Z}_1, \dots, \mathbf{Z}_p$ *be* $p+1$ *statistics with* $\mathrm{E}(\mathbf{T}) = \boldsymbol{\theta} \in \mathbb{R}^l$ *and* $\mathrm{E}(\mathbf{Z}_i) = \mathbf{0} \in \mathbb{R}^l$ *for* $i = 1, \dots, p$. *Further let*

$$\begin{aligned} \mathbf{g} &= (\mathrm{tr}[\mathrm{Cov}(\mathbf{T}, \mathbf{Z}_1)], \dots, \mathrm{tr}[\mathrm{Cov}(\mathbf{T}, \mathbf{Z}_p)])^{\mathrm{T}} \quad \text{and} \\ \mathbf{G} &= \begin{pmatrix} \mathrm{tr}[\mathrm{Cov}(\mathbf{Z}_1)] & \mathrm{tr}[\mathrm{Cov}(\mathbf{Z}_1, \mathbf{Z}_2)] & \dots & \mathrm{tr}[\mathrm{Cov}(\mathbf{Z}_1, \mathbf{Z}_p)] \\ \mathrm{tr}[\mathrm{Cov}(\mathbf{Z}_2, \mathbf{Z}_1)] & \mathrm{tr}[\mathrm{Cov}(\mathbf{Z}_2)] & \dots & \mathrm{tr}[\mathrm{Cov}(\mathbf{Z}_2, \mathbf{Z}_p)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}[\mathrm{Cov}(\mathbf{Z}_p, \mathbf{Z}_1)] & \mathrm{tr}[\mathrm{Cov}(\mathbf{Z}_p, \mathbf{Z}_2)] & \dots & \mathrm{tr}[\mathrm{Cov}(\mathbf{Z}_p)] \end{pmatrix} \end{aligned}$$

be known quantities where $\mathbf{G}$ *is a positive definite matrix.*
For $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)^{\mathrm{T}}$ *we consider estimators of the type*

$$\mathbf{T}_{\boldsymbol{\gamma}} = \mathbf{T} - \sum_{i=1}^{p} \gamma_i \mathbf{Z}_i$$

which are unbiased for $\boldsymbol{\theta}$.
Then the total variance $\mathrm{tr}[\mathrm{Cov}(\mathbf{T}_{\boldsymbol{\gamma}})]$ *of* $\mathbf{T}_{\boldsymbol{\gamma}}$ *is minimized by the choice*

$$\boldsymbol{\gamma}_{\mathrm{opt}} = \mathbf{G}^{-1}\mathbf{g}$$

and the corresponding minimal value of the total variance is given by

$$\mathrm{tr}[\mathrm{Cov}(\mathbf{T}_{\boldsymbol{\gamma}_{\mathrm{opt}}})] = \mathrm{tr}[\mathrm{Cov}(\mathbf{T})] - \mathbf{g}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{g}\,.$$

The final lemma in this section is concerned with the positive definiteness of certain submatrices of a positive definite matrix $\mathbf{A}$.

Lemma A.9 *Let the partitioned matrix*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \dots & \mathbf{A}_{1k} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \dots & \mathbf{A}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{k1} & \mathbf{A}_{k2} & \dots & \mathbf{A}_{kk} \end{pmatrix} \in \mathbb{R}^{kl \times kl} \tag{A.1}$$

consisting of $l \times l$-dimensional blocks be positive definite. Then the matrices

$$\mathbf{A}_j = \begin{pmatrix} A_{11,jj} & A_{12,jj} & \dots & A_{1k,jj} \\ A_{21,jj} & A_{22,jj} & \dots & A_{2k,jj} \\ \vdots & \vdots & \ddots & \vdots \\ A_{k1,jj} & A_{k2,jj} & \dots & A_{kk,jj} \end{pmatrix} \in \mathbb{R}^{k \times k} \tag{A.2}$$

consisting of the j-th main diagonal elements of the blocks, for $j = 1, \dots, l$, and

$$\mathbf{A}_{\text{tr}} = \begin{pmatrix} \text{tr}(\mathbf{A}_{11}) & \text{tr}(\mathbf{A}_{12}) & \dots & \text{tr}(\mathbf{A}_{1k}) \\ \text{tr}(\mathbf{A}_{21}) & \text{tr}(\mathbf{A}_{22}) & \dots & \text{tr}(\mathbf{A}_{2k}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{tr}(\mathbf{A}_{k1}) & \text{tr}(\mathbf{A}_{k2}) & \dots & \text{tr}(\mathbf{A}_{kk}) \end{pmatrix} \in \mathbb{R}^{k \times k} \tag{A.3}$$

consisting of the traces of the blocks are positive definite as well.

Proof:
A *submatrix* of a matrix $\mathbf{A}$ is a matrix that can be obtained by striking out rows and/or columns of $\mathbf{A}$. A submatrix of a quadratic matrix is called a *principal submatrix* if it can be obtained by striking out the same rows as columns (so that the i-th row is struck out whenever the i-th column is struck out, and vice versa); cf. HARVILLE (1997), pp. 13–14. The $\mathbf{A}_j$ are principle submatrices obtained by striking out all rows and columns except for the j-th, $(l+j)$-th, $(2l+j)$-th, ... , $((k-1)l+j)$-th.
According to HARVILLE (1997), Corollary 14.2.12, any principal submatrix of a positive definite matrix is positive definite, such that $\mathbf{A}_j$ is positive definite for $j = 1, \dots, l$.
Since

$$\mathbf{A}_{\text{tr}} = \sum_{j=1}^{l} \mathbf{A}_j \tag{A.4}$$

the positive definiteness of $\mathbf{A}_{\text{tr}}$ follows. ∎

B Proof of Lemma 3.1

Lemma 3.1 *The optimal parameter matrix* $\mathbf{B}_{\text{opt}}$ *for the combination* $\mathbf{f}_{\mathbf{B}_{\text{opt}},\text{rest}}$ *is not altered if we use any other constant vector than* $\boldsymbol{\mu}_0$ *in the calculation of the covariance matrix*

$$\breve{\mathbf{\Sigma}} = \mathrm{E}\left[\left(\begin{pmatrix} \mathbf{y} \\ \mathbf{f} \end{pmatrix} - (\mathbf{1}_{k+1} \otimes \boldsymbol{\mu}_0)\right)\left(\begin{pmatrix} \mathbf{y} \\ \mathbf{f} \end{pmatrix} - (\mathbf{1}_{k+1} \otimes \boldsymbol{\mu}_0)\right)^{\mathrm{T}}\right] .$$

Proof: We show that the optimal parameter matrix $\mathbf{B}_{\text{opt}}$ is the same regardless whether we use

$$\breve{\mathbf{\Sigma}} = \mathrm{E}\left[\left(\begin{pmatrix} \mathbf{y} \\ \mathbf{f} \end{pmatrix} - (\mathbf{1}_{k+1} \otimes \boldsymbol{\mu}_0)\right)\left(\begin{pmatrix} \mathbf{y} \\ \mathbf{f} \end{pmatrix} - (\mathbf{1}_{k+1} \otimes \boldsymbol{\mu}_0)\right)^{\mathrm{T}}\right]$$

or

$$\tilde{\mathbf{\Sigma}} = \mathrm{E}\left[\left(\begin{pmatrix} \mathbf{y} \\ \mathbf{f} \end{pmatrix} - (\mathbf{1}_{k+1} \otimes \boldsymbol{\nu}_0)\right)\left(\begin{pmatrix} \mathbf{y} \\ \mathbf{f} \end{pmatrix} - (\mathbf{1}_{k+1} \otimes \boldsymbol{\nu}_0)\right)^{\mathrm{T}}\right] .$$

in its calculation, where $\boldsymbol{\nu}_0 \in \mathbb{R}^l$ is arbitrary.
From Equations (3.17) and (3.4) we may conclude that the target function $\mathrm{MMSPE}(\mathbf{f}_{\mathbf{B},\text{rest}}, \mathbf{y})$ depends on the moments of the joint distribution of $(\mathbf{y}^{\mathrm{T}}, \mathbf{f}^{\mathrm{T}})^{\mathrm{T}}$ only via the moments of the errors $\mathbf{e} = (\mathbf{e}_1^{\mathrm{T}}, \ldots, \mathbf{e}_k^{\mathrm{T}})^{\mathrm{T}}$. Since $\mathrm{E}(\mathbf{e}_i) = \mathbf{0}$ for $i = 1, \ldots, k$ because of the unbiasedness assumption, it suffices to show that $\breve{\mathbf{V}} = \mathrm{E}(\mathbf{e}\mathbf{e}^{\mathrm{T}})$ is independent of the choice of $\boldsymbol{\nu}_0$. We will show that each submatrix $\breve{\mathbf{V}}_{ij}$ is independent of the choice of $\boldsymbol{\nu}_0$:

$$\begin{aligned} \breve{\mathbf{V}}_{ij} &= \mathrm{E}[\mathbf{e}_i\mathbf{e}_j^{\mathrm{T}}] = \mathrm{E}[(\mathbf{f}_i - \mathbf{y})(\mathbf{f}_j - \mathbf{y})^{\mathrm{T}}] \\ &= \mathrm{E}[((\mathbf{f}_i - \boldsymbol{\nu}) - (\mathbf{y} - \boldsymbol{\nu}))((\mathbf{f}_j - \boldsymbol{\nu}) - (\mathbf{y} - \boldsymbol{\nu}))^{\mathrm{T}}] \\ &= \mathrm{E}[(\mathbf{f}_i - \boldsymbol{\nu})(\mathbf{f}_j - \boldsymbol{\nu})^{\mathrm{T}}] - \mathrm{E}[(\mathbf{f}_i - \boldsymbol{\nu})(\mathbf{y} - \boldsymbol{\nu})^{\mathrm{T}}] \\ &\quad - \mathrm{E}[(\mathbf{y} - \boldsymbol{\nu})(\mathbf{f}_j - \boldsymbol{\nu})^{\mathrm{T}}] + \mathrm{E}[(\mathbf{y} - \boldsymbol{\nu})(\mathbf{y} - \boldsymbol{\nu})^{\mathrm{T}}] , \end{aligned} \tag{B.1}$$

where $\boldsymbol{\nu} \in \mathbb{R}^l$ is arbitrary. If we set $\boldsymbol{\nu} = \boldsymbol{\mu}_0$ in the final expression we obtain

$$\breve{\mathbf{V}}_{ij} = \breve{\mathbf{\Sigma}}_{ij} - \breve{\mathbf{\Sigma}}_{i0} - \breve{\mathbf{\Sigma}}_{0j} + \breve{\mathbf{\Sigma}}_{00} , \tag{B.2}$$

whereas for $\boldsymbol{\nu} = \boldsymbol{\nu}_0$ we obtain

$$\breve{\mathbf{V}}_{ij} = \tilde{\mathbf{\Sigma}}_{ij} - \tilde{\mathbf{\Sigma}}_{i0} - \tilde{\mathbf{\Sigma}}_{0j} + \tilde{\mathbf{\Sigma}}_{00} . \tag{B.3}$$

This completes the proof. ■

Acknowledgements: The author wishes to thank Götz Trenkler and Jürgen Groß for their helpful comments and suggestions. The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

BAKSALARY, J.K. and KALA, R. (1982): 'Admissible estimation by covariance adjustment technique'. *Sankhya Series A* **44**, 281-285.

CLEMEN, R.T. (1989): 'Combining forecasts: A review and annotated bibliography'. *International Journal of Forecasting* **5**, 559-583.

FUHRER, J. and HALTMAIER, J. (1988): 'Minimum variance pooling of forecasts at different levels of aggregation'. *Journal of Forecasting* **7**, 63-73.

GRANGER, C.W.J. (1989): 'Combining forecasts – Twenty years later'. *Journal of Forecasting* **8**, 167-173.

GRANGER, C.W.J. and RAMANATHAN, R. (1984): 'Improved methods of combining forecasts'. *Journal of Forecasting* **3**, 197-204.

HARVILLE, D.A. (1985): 'Decomposition of prediction error'. *Journal of the American Statistical Association* **80**, 132-138.

HARVILLE, D.A. (1997): Matrix Algebra From a Statistician's Perspective. *Springer, New York.*

IHORST, G. (1993): Verbesserte Schätzverfahren auf der Grundlage des Covariance-Adjustment-Prinzips. *Verlag Anton Hain, Frankfurt am Main.*

KANG, H. (1986): 'Unstable weights in the combination of forecasts'. *Management Science* **32**, 683-695.

KLAPPER, M. (1999): Multivariate Rank-Based Forecast Combining Techniques. *Technical Report 2/1999, Sonderforschungsbereich 475, University of Dortmund.*

KLAPPER, M. (2000): The Combination of Forecasts Using Rank-Based Techniques. *Josef Eul Verlag, Lohmar.*

LÖWNER, K. (1934): 'Über monotone Matrixfunktionen'. *Mathematische Zeitschrift* **38**, 177-216.

MAGNUS, J.R. and NEUDECKER, H. (1999): Matrix Differential Calculus with Applications in Statistics and Econometrics, Revised Edition. *Wiley, Chichester.*

ODELL, P.L., DORSETT, D., YOUNG, D. and IGWE, J. (1989): 'Estimator models for combining vector estimators'. *Mathematical and Computer Modeling* **12**, 1627-1642.

RAO, A.R. and BHIMASANKARAM, P. (1992): Linear Algebra. *Tata McGraw-Hill, New Delhi.*

RAO, C.R. (1966): 'Covariance adjustment and related problems in multivariate analysis'. In: KRISHNAIAH, P.R. (ed.), Multivariate Analysis. *Academic Press, New York*, 87-103.

RAO, C.R. (1967): 'Least squares theory using an estimated dispersion matrix and its application to measurement of signals'. In: LECAM, L.M. and NEYMAN, J. (eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1. *University of California Press, Berkeley, CA*, 355-372.

THIELE, J. (1993): Kombination von Prognosen. (Wirtschaftswissenschaftliche Beiträge: Band 74). *Physica, Heidelberg.*

TRENKLER, G. and IHORST, G. (1995): 'Improved estimation by weak covariance adjustment technique'. *Discussiones Mathematicae Algebra and Stochastic Methods* **15**, 189-201.

TROSCHKE, S.O. (2002): Regression Approach to the Linear Combination of Multivariate Forecasts. *Technical Report 4/2002, Sonderforschungsbereich 475, University of Dortmund.*

TROSCHKE, S.O. and TRENKLER, G. (2000): Linear Plus Quadratic Approach to the Mean Square Error Optimal Combination of Forecasts. *Technical Report 54/2000, Sonderforschungsbereich 475, University of Dortmund.*

WENZEL, T. (1998): Pitman-Closeness and the Linear Combination of Multivariate Forecasts. *Technical Report 34/1998, Sonderforschungsbereich 475, University of Dortmund.*

WENZEL, T. (1999a): Using Different Pitman-Closeness Techniques for the Linear Combination of Multivariate Forecasts. *Technical Report 18/1999, Sonderforschungsbereich 475, University of Dortmund.*

WENZEL, T. (1999b): Combination of Biased Forecasts: Bias Correction or Bias Based Weights. *Technical Report 50/1999, Sonderforschungsbereich 475, University of Dortmund.*

WENZEL, T. (2000): The Shrinkage Approach in the Combination of Forecasts. *Technical Report 44/2000, Sonderforschungsbereich 475, University of Dortmund.*

WENZEL, T. (2001): Die Kombination von Prognosen unter Berücksichtigung statistischer Bewertungskriterien. *Josef Eul Verlag, Lohmar.*