Regression Approach to the Linear Plus Quadratic Combination of Forecasts

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

Abstract: In Troschke and Trenkler (2000) the authors introduce linear plus quadratic approaches to the mean square error optimal combination of forecasts for a scalar random variable. In this paper it is shown how the optimal combination parameters can be obtained with the help of linear regression. Thus numerical considerations as well as application of linear plus quadratic combination to empirical data are facilitated. First results on the comparison of the new methods to the classical linear approaches are given. It is found that there are situations where the linear plus quadratic approaches may be employed beneficially, but further investigations have to be carried out.

Keywords: Combination of forecasts, linear plus quadratic combination, linear regression.

AMS 2000 Subject Classification: 62M20, 62J05, 62P20

1 Introduction

Suppose that we are given k forecasts $f_1, \ldots, f_k$ for a scalar random variable y. The forecasts are gathered in a random vector $\mathbf{f}$, i.e. $\mathbf{f} = (f_1, \ldots, f_k)^{\mathrm{T}}$. It is a common procedure to combine the individual forecasts f_i, in order to obtain a single improved forecast for the target variable y.

In this paper we regard improvement with respect to the mean square prediction error

$$\mathrm{MSPE}(f, y) = \mathrm{E}[(y - f)^2] \tag{1.1}$$

of a forecast f for a target variable y.

Linear combinations have been used predominantly for that purpose, compare e.g. CLEMEN (1989) or THIELE (1993) for good overviews on the topic. Linear forecast combinations are of the form $\mathbf{b}^{\mathrm{T}}\mathbf{f} + c$ with $c \in \mathbb{R}$ and $\mathbf{b} = (b_1, \ldots, b_k)^{\mathrm{T}} \in \mathbb{R}^k$, where it may be appropriate to impose certain restrictions on the combination parameters $\mathbf{b}$ and c. It is well-known (see e.g. THIELE, 1993) that a linear combination $f_{\mathbf{b},c} = \mathbf{b}^{\mathrm{T}}\mathbf{f} + c$ with suitably chosen $\mathbf{b} = (b_1, \ldots, b_k)^{\mathrm{T}} \in \mathbb{R}^k$ and $c \in \mathbb{R}$ is optimal among *all* combinations if y and $\mathbf{f}$ follow a joint normal distribution.

In the absence of joint normality, however, it is worthwhile to consider nonlinear forecast combinations. Stimulated by TAYLOR's series expansion formula TROSCHKE and TRENKLER (2000) introduce linear plus quadratic combinations of the form $\mathbf{f}^{\mathrm{T}}\mathbf{A}\mathbf{f} + \mathbf{b}^{\mathrm{T}}\mathbf{f} + c$, where $\mathbf{A}$ is a $k \times k$ real symmetric matrix, $\mathbf{b} = (b_1, \ldots, b_k)^{\mathrm{T}} \in \mathbb{R}^k$ and $c \in \mathbb{R}$. Again, restricting the combination parameters may be reasonable. We refer to Section 3 for details on the employed restrictions and the classes of linear plus quadratic combinations evolving from them.

In their paper the authors show how the combination parameters should be chosen within the corresponding classes in order to minimize the mean square prediction error of the combined forecast. The optimal combination parameters depend on the first to fourth order moments of the joint distribution of the target variable and its forecasts. The results are compared to those for several linear combinations, where only the first and second order moments are involved. In practical applications such moments will hardly be known and the authors suggest to estimate the necessary moments from a sample of observations on the variables of interest.

For some linear plus quadratic combinations the optimal parameters cannot be given explicitly but only in terms of a complicated linear equation system. Consequently, also the optimal MSPE-value which may be achieved within the considered class of linear plus quadratic combinations cannot be given by an explicit expression. Even more important, application of linear plus quadratic combination is impeded, since only those numbers k of forecasts can be dealt with for which the equation system has been made explicit. Even for the case of $k = 2$ forecasts this turns out to be a cumbersome task. Consequently, it is desirable to find an easier way to apply linear plus quadratic combination. This easier approach is presented in the following: Similar to the findings of GRANGER and RAMANATHAN (1984) for linear combinations we introduce a linear regression approach for linear plus quadratic combinations. Thus it is straightforward to implement linear plus quadratic combination for any number k of forecasts and standard computer software becomes applicable.

While the linear regression approach does not further theoretical insights on linear plus quadratic combination, it is most helpful in achieving *two* practical *goals*: On the one hand, based on given first to fourth order moments of the joint distribution of y and $\mathbf{f}$ we want to calculate which optimal MSPE-values may be reached in that situation. This allows for numerical comparisons of the potential inherent in different combination techniques. On the other hand, we want to facilitate application of the combination techniques to empirical data. By employing the linear regression approach both is possible for any number k of forecasts without additional effort. Thus we can also carry out investigations on the appropriate choice of k easily.

In TROSCHKE and TRENKLER (2000) also the case of $k = 1$ forecast is investigated, which results in linear and linear plus quadratic adjustments of single forecasts. The linear regression approach covers these adjustments as well.

Linear and linear plus quadratic adjustments and combinations of forecasts will be compared in a small numerical example with respect to their potential as well as with respect to their performance for empirical data (see Section 4).

In our derivations we will have to consider the first to fourth order moments of the joint distribution of y and $\mathbf{f}$. The following notations will be useful:

Extending the approach from HARVILLE (1985) and utilizing the notations from RAO and KLEFFE (1988) we will assume the following setting: The expectations of y and $\mathbf{f}$ are given by $\mathrm{E}(y) = \mu_0$ and $\mathrm{E}(\mathbf{f}) = \boldsymbol{\mu}_\mathbf{f} := (\mu_1, \ldots, \mu_k)^\mathrm{T}$, respectively, which gives rise to the model:

$$\begin{pmatrix} y \\ \mathbf{f} \end{pmatrix} = \begin{pmatrix} \mu_0 \\ \boldsymbol{\mu}_\mathbf{f} \end{pmatrix} + \begin{pmatrix} \varepsilon_0 \\ \boldsymbol{\varepsilon}_\mathbf{f} \end{pmatrix} =: \boldsymbol{\mu} + \boldsymbol{\varepsilon} , \tag{1.2}$$

where $\boldsymbol{\varepsilon}_\mathbf{f} := (\varepsilon_1, \ldots, \varepsilon_k)^\mathrm{T}$. Consequently, $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and the higher order moments of $\boldsymbol{\varepsilon}$ are the centered moments of $(y, \mathbf{f}^\mathrm{T})^\mathrm{T}$.

First, let us turn to the second order moments:

$$\boldsymbol{\Sigma} := \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\mathrm{T}) = \mathrm{E}\left[\begin{pmatrix} \varepsilon_0 \\ \boldsymbol{\varepsilon}_\mathbf{f} \end{pmatrix} \begin{pmatrix} \varepsilon_0 \\ \boldsymbol{\varepsilon}_\mathbf{f} \end{pmatrix}^\mathrm{T} \right] =: \begin{pmatrix} \Sigma_{00} & \boldsymbol{\Sigma}_{0\mathbf{f}} \\ \boldsymbol{\Sigma}_{\mathbf{f}0} & \boldsymbol{\Sigma}_{\mathbf{ff}} \end{pmatrix} \tag{1.3}$$

and

$$\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\mathrm{T}) = \mathrm{E}\left[\left(\begin{pmatrix} y \\ \mathbf{f} \end{pmatrix} - \begin{pmatrix} \mu_0 \\ \boldsymbol{\mu}_\mathbf{f} \end{pmatrix} \right) \left(\begin{pmatrix} y \\ \mathbf{f} \end{pmatrix} - \begin{pmatrix} \mu_0 \\ \boldsymbol{\mu}_\mathbf{f} \end{pmatrix} \right)^\mathrm{T} \right] = \mathrm{Cov} \begin{pmatrix} y \\ \mathbf{f} \end{pmatrix} . \tag{1.4}$$

The lower left $(k \times 1)$–submatrix $\boldsymbol{\Sigma}_{\mathbf{f}0}$ and the lower right $(k \times k)$–submatrix $\boldsymbol{\Sigma}_{\mathbf{ff}}$ of

$\boldsymbol{\Sigma}$ read explicitly

$$\boldsymbol{\Sigma}_{\mathbf{f}0} = \begin{pmatrix} \Sigma_{10} \\ \Sigma_{20} \\ \vdots \\ \Sigma_{k0} \end{pmatrix} \quad \text{and } \boldsymbol{\Sigma}_{\mathbf{ff}} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1k} \\ \Sigma_{21} & \Sigma_{22} & \dots & \Sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{k1} & \Sigma_{k2} & \dots & \Sigma_{kk} \end{pmatrix} . \tag{1.5}$$

Note that vectors and matrices are represented by bold face letters. Analogously, the third order moments of $\boldsymbol{\varepsilon}$ are given by

$$\boldsymbol{\Phi} := \mathrm{E}(\boldsymbol{\varepsilon} \otimes \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = \begin{pmatrix} \boldsymbol{\Phi}_0 \\ \boldsymbol{\Phi}_1 \\ \vdots \\ \boldsymbol{\Phi}_k \end{pmatrix} , \tag{1.6}$$

where

$$\boldsymbol{\Phi}_i = \mathrm{E}(\varepsilon_i \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = \begin{pmatrix} \Phi_{i00} & \boldsymbol{\Phi}_{i0\mathbf{f}} \\ \boldsymbol{\Phi}_{i\mathbf{f}0} & \boldsymbol{\Phi}_{i\mathbf{ff}} \end{pmatrix} , \; i = 0 , \dots , k \tag{1.7}$$

and the fourth order moments are given by

$$\boldsymbol{\Psi} = \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}} \otimes \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = \begin{pmatrix} \boldsymbol{\Psi}_{00} & \boldsymbol{\Psi}_{01} & \dots & \boldsymbol{\Psi}_{0k} \\ \boldsymbol{\Psi}_{10} & \boldsymbol{\Psi}_{11} & \dots & \boldsymbol{\Psi}_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{k0} & \boldsymbol{\Psi}_{k1} & \dots & \boldsymbol{\Psi}_{kk} \end{pmatrix} , \tag{1.8}$$

where

$$\boldsymbol{\Psi}_{ij} = \mathrm{E}(\varepsilon_i \varepsilon_j \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = \begin{pmatrix} \Psi_{ij00} & \boldsymbol{\Psi}_{ij0\mathbf{f}} \\ \boldsymbol{\Psi}_{ij\mathbf{f}0} & \boldsymbol{\Psi}_{ij\mathbf{ff}} \end{pmatrix} , \; i, j = 0 , \dots , k. \tag{1.9}$$

The elements of $\boldsymbol{\Phi}$ are $\Phi_{ijl} = \mathrm{E}(\varepsilon_i \varepsilon_j \varepsilon_l)$ and the elements of $\boldsymbol{\Psi}$ are $\Psi_{ijlm} = \mathrm{E}(\varepsilon_i \varepsilon_j \varepsilon_l \varepsilon_m)$.

Section 2 resumes the classical linear combinations and their representation in terms of linear regression. In Section 3 the respective linear plus quadratic combinations are investigated. Two facets of a linear regression representation are derived to achieve the above defined goals. The already mentioned numerical example is reported in Section 4. Section 5 concludes the paper. Finally, Section A in the appendix lists two results which are fundamental for our derivations.

2 Linear combinations

We will consider four versions of the linear combination approach $\mathbf{b}^{\mathrm{T}}\mathbf{f} + c$ which vary with respect to the restrictions imposed on the combination parameters $\mathbf{b}$ and c. Granger and Ramanathan (1984) showed that these linear combinations are closely related to certain linear regression models.
The first version is

$$f_{\mathbf{b},c} = \mathbf{b}^{\mathrm{T}}\mathbf{f} + c \; . \tag{2.1}$$

As stated in Section 1, with suitably chosen parameters, this version leads to the MSPE-optimal combined forecast under joint normality of y and $\mathbf{f}$. Following Granger and Ramanathan (1984) the optimal combination parameters may be obtained by regressing the target variable y on the individual forecasts $\mathbf{f}$, using a constant term in the regression.
A simpler approach is to define the combined forecast to be a weighted average of the single forecasts

$$f_{\mathbf{b}} = \mathbf{b}^{\mathrm{T}}\mathbf{f} \; . \tag{2.2}$$

This corresponds to regressing y on $\mathbf{f}$ without a constant term. Clearly, from the standpoint of regression analysis the goodness of fit decreases by dropping the constant term, but sometimes empirical combination results improve by doing so.
If each of the single forecasts is unbiased it is a well-known fact that the combined forecast is unbiased as well if, in the second approach, the parameters are chosen such that they sum up to unity, i.e. $\mathbf{b}^{\mathrm{T}}\mathbf{1} = 1$. This leads to the third version of the linear approach which utilizes this restriction:

$$f_{\mathbf{b},\mathrm{rest}} = \mathbf{b}^{\mathrm{T}}\mathbf{f} \; , \text{ where } \mathbf{b}^{\mathrm{T}}\mathbf{1} = 1 \; . \tag{2.3}$$

The corresponding regression model is to regress $y - f_1$ on $f_2 - f_1, \ldots, f_k - f_1$ without a constant term. Thus the parameters $b_2, \ldots, b_k$ are obtained while b_1 results from $b_1 = 1 - \sum_{i=2}^{k} b_i$.
If the individual forecasts f_i are biased it is reasonable to perform a bias correction $f_i - \mu_i + \mu_0$ before combining them. After the correction the individual forecasts are unbiased and, hence, they should be combined with weights summing up to unity. This leads to the restricted linear combination with absolute term:

$$f_{\mathbf{b},c,\mathrm{rest}} = \mathbf{b}^{\mathrm{T}}\mathbf{f} + c \; , \text{ where } \mathbf{b}^{\mathrm{T}}\mathbf{1} = 1 \; . \tag{2.4}$$

Correspondingly, $y - f_1$ should be regressed on $f_2 - f_1, \ldots, f_k - f_1$, using a constant term. Again b_1 is obtained from $b_1 = 1 - \sum_{i=2}^{k} b_i$.

Clearly, from the point of view of regression analysis the unrestricted combination with constant term $f_{\mathbf{b},c}$ provides the best fit in general. The other combinations are appropriate only if the restrictions seem justified in the situation under consideration. The popular restricted combined forecast without constant term $f_{\mathbf{b},\text{rest}}$, for example, is advantageous, if all single forecasts are unbiased as was mentioned above. The assumption of unbiasedness for each single forecast, however, seems at least doubtful. GRANGER and RAMANATHAN (1984, p. 200) point out:

> There is nothing sacred about the weights adding up to unity, although that seems to be the common practice. Furthermore, there is no reason to believe that every alternative forecast will be unbiased.

Our exposition of important linear combinations would not be complete without the arithmetic mean of the individual forecasts:

$$f_{\text{am}} = \frac{1}{k}\sum_{i=1}^{k} f_i = \frac{1}{k}\mathbf{1}^{\mathrm{T}}\mathbf{f} \ . \tag{2.5}$$

Here no regression is necessary, since the combination parameters are fixed as $b_i = 1/k$, $i = 1, \ldots, k$ and $c = 0$. Nevertheless this simple combination proves to be very powerful in empirical applications.

If we consider the special case $k = 1$ we arrive at *adjustments of individual forecasts* f_i. The performance of f_i can be improved by this kind of adjustment. All of the linear combination approaches described above may be employed in this case. Some of them, however, are identical to others, as we will see in the following:

The *unrestricted linear adjustment with constant term* is

$$(f_i)_{b,c} = bf_i + c \tag{2.6}$$

with $b, c \in \mathbb{R}$. GRANGER (1989, p. 169) points out the usefulness of such an adjustment. The *unrestricted linear adjustment without constant term* reads

$$(f_i)_b = bf_i \tag{2.7}$$

with $b \in \mathbb{R}$. The *linear adjustment with constant term and with the restriction of the weights summing up to unity* is

$$(f_i)_{1,c} = f_i + c \ . \tag{2.8}$$

The optimal choice for $c \in \mathbb{R}$ results in the well known bias corrected forecast. Finally, the *linear adjustment without constant term and with the restriction of the weights summing up to unity* as well as the adjustment counterpart of the *arithmetic mean* equal the original single forecast f_i and need no special consideration. The regression models corresponding to the adjustments are obvious from the above exposition for $k \geq 2$ forecasts.
We now turn to the linear plus quadratic approaches to the combination of forecasts. Since the linear combination $f_{\mathbf{b},c} = \mathbf{b}^{\mathrm{T}}\mathbf{f} + c$ with appropriately chosen weights is MSPE-optimal among *all* combined forecasts under joint normality of y and $\mathbf{f}$, employment of linear plus quadratic approaches only deserves attention under non-normality. Hence we will assume non-normality in the following.

3 Linear plus quadratic combinations

Linear plus quadratic combinations are of the general form $\mathbf{f}^{\mathrm{T}}\mathbf{A}\mathbf{f} + \mathbf{b}^{\mathrm{T}}\mathbf{f} + c$, where $c \in \mathbb{R}$, $\mathbf{b} = (b_1, \ldots, b_k)^{\mathrm{T}} \in \mathbb{R}^k$ and

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ a_{12} & a_{22} & \ldots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k} & a_{2k} & \ldots & a_{kk} \end{pmatrix} \in \mathbb{R}^{k \times k} \tag{3.1}$$

is a symmetric matrix. The versions analyzed here have been introduced in Troschke and Trenkler (2000). They differ with respect to the choice of the symmetric matrix $\mathbf{A}$ in the quadratic part:
The *strong* version

$$f_{\mathbf{A},\mathbf{b},c} = \mathbf{f}^{\mathrm{T}}\mathbf{A}\mathbf{f} + \mathbf{b}^{\mathrm{T}}\mathbf{f} + c \tag{3.2}$$

uses a full matrix $\mathbf{A}$, the *medium* version

$$f_{\mathbf{a},\mathbf{b},c} = \mathbf{f}^{\mathrm{T}}\,\mathrm{dg}(\mathbf{a})\mathbf{f} + \mathbf{b}^{\mathrm{T}}\mathbf{f} + c = \sum_{i=1}^{k} a_i f_i^2 + \mathbf{b}^{\mathrm{T}}\mathbf{f} + c \tag{3.3}$$

a diagonal matrix

$$\mathbf{A} = \mathrm{dg}(\mathbf{a}) = \begin{pmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_k \end{pmatrix} \in \mathbb{R}^{k \times k}\ , \tag{3.4}$$

where $\mathbf{a} = (a_1, \dots, a_k)^{\mathrm{T}} \in \mathbb{R}^k$, and the *weak* version

$$f_{\alpha,\mathbf{b},c} = \alpha \mathbf{f}^{\mathrm{T}}\mathbf{f} + \mathbf{b}^{\mathrm{T}}\mathbf{f} + c \tag{3.5}$$

uses a multiple of the $k \times k$-identity matrix $\mathbf{A} = \alpha\mathbf{I}$.
The respective choices of the matrix in the quadratic part may be viewed as restrictions on $\mathbf{A}$ with the effect that the number of parameters involved is reduced from $(k+1)(k+2)/2$ over $2k+1$ to $k+2$. Since the number of observations from which the unknown parameters are to be estimated in empirical applications is not so large in general, this reduction of the number of parameters may be reasonable.

3.1 Strong linear plus quadratic combination

In order to facilitate numerical considerations as well as application of linear plus quadratic forecast combinations it is important to note, that we may regard the problem of finding the optimal combination parameters as a linear regression problem just like it is the case with the linear combination approaches (cf. Section 2).
We rewrite the MSPE-function belonging to the strong linear plus quadratic combination with the help of Lemma A.1:

$$\begin{aligned} \mathrm{MSPE}(f_{\mathbf{A},\mathbf{b},c}, y) &= \mathrm{E}[(y - f_{\mathbf{A},\mathbf{b},c})^2] = \mathrm{E}[(y - \mathbf{f}^{\mathrm{T}}\mathbf{A}\mathbf{f} - \mathbf{b}^{\mathrm{T}}\mathbf{f} - c)^2] \\ &= \mathrm{E}\left[\left(y - \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij} f_i f_j - \sum_{i=1}^{k} b_i f_i - c\right)^2\right] \\ &= \mathrm{E}\left[\left(y - \sum_{i=1}^{k} a_{ii} f_i^2 - 2\sum_{i}\sum_{<\, j} a_{ij} f_i f_j - \sum_{i=1}^{k} b_i f_i - c\right)^2\right] . \end{aligned} \tag{3.6}$$

Minimization of this function corresponds to the linear regression problem of regressing the target variable y on the vector $\mathbf{g} = (g_1, \dots, g_{k(k+1)/2+k})^{\mathrm{T}} = (f_1^2, \dots, f_k^2, (f_i f_j)_{i,j=1,\dots,k, i<j}, f_1, \dots, f_k)^{\mathrm{T}}$, i.e. on the vector of squared forecasts f_i^2, mixed products $f_i f_j$ and forecasts f_i, using a constant term, cf. RAO (1965, pp. 222 f.).
The coefficients $(\omega_0, \boldsymbol{\omega}^{\mathrm{T}})^{\mathrm{T}} = (\omega_0, \omega_1, \dots, \omega_{k(k+1)/2+k})^{\mathrm{T}}$ obtained by this regression are the combination parameters: While $\omega_0 = c$, the vector $\boldsymbol{\omega}$ consists of the elements a_{ii}, $i = 1, \dots, k$, $2a_{ij}$, $i, j = 1, \dots, k$, $i < j$ (notice the factor 2 premultiplying a_{ij}) and b_i, $i = 1, \dots, k$.
Following RAO (1965) these regression coefficients, and hence the theoretically optimal choices $\mathbf{A}_{\mathrm{opt}}$, $\mathbf{b}_{\mathrm{opt}}$ and c_{opt} based on given moments $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, can be

calculated as follows: The vector $\boldsymbol{\omega}$ is given by

$$\boldsymbol{\omega} = \mathrm{Cov}(\mathbf{g})^{-1}\,\mathrm{Cov}(\mathbf{g}, y)\ , \tag{3.7}$$

where $\mathrm{Cov}(\mathbf{g})$ is the covariance matrix of regression variables, while $\mathrm{Cov}(\mathbf{g}, y)$ is the vector of covariances between regression variables and target variable. The constant term ω_0 is given by

$$\omega_0 = \mathrm{E}(y) - \boldsymbol{\omega}^{\mathrm{T}}\,\mathrm{E}(\mathbf{g})\ . \tag{3.8}$$

The necessary covariances and expectations may be obtained from Lemma A.2 under the setting

$$\tilde{\mathbf{Y}} = \begin{pmatrix} y \\ \mathbf{f} \end{pmatrix}\ ,\ \tilde{\boldsymbol{\mu}} = \begin{pmatrix} \mu_0 \\ \boldsymbol{\mu}_{\mathbf{f}} \end{pmatrix} = \boldsymbol{\mu} \text{ and } \tilde{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon_0 \\ \boldsymbol{\varepsilon}_{\mathbf{f}} \end{pmatrix} = \boldsymbol{\varepsilon} \tag{3.9}$$

leading to

$$\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}\ ,\ \tilde{\boldsymbol{\Phi}} = \boldsymbol{\Phi} \text{ and } \tilde{\boldsymbol{\Psi}} = \boldsymbol{\Psi} \tag{3.10}$$

as defined in (1.3) and (1.6) – (1.9). We observe that for $i, j = 1, \ldots, k$

$$\begin{aligned} f_i f_j &= \tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{A}}\tilde{\mathbf{Y}} \quad \text{with} \quad \tilde{\mathbf{A}} = \frac{1}{2}(\mathbf{E}_{ij} + \mathbf{E}_{ji}) \text{ symmetric} \\ f_i &= \tilde{\mathbf{b}}^{\mathrm{T}}\tilde{\mathbf{Y}} \quad \text{with} \quad \tilde{\mathbf{b}} = \boldsymbol{\xi}_i^{(k+1)} \\ y &= \tilde{\mathbf{b}}^{\mathrm{T}}\tilde{\mathbf{Y}} \quad \text{with} \quad \tilde{\mathbf{b}} = \boldsymbol{\xi}_0^{(k+1)}\ . \end{aligned} \tag{3.11}$$

Here $\mathbf{E}_{ij}$ denotes the $(k+1) \times (k+1)$ matrix with elements $\tilde{a}_{lm}$, $l, m = 0, \ldots, k$ where $\tilde{a}_{lm} = 1$ if $(l, m) = (i, j)$ and $\tilde{a}_{lm} = 0$ otherwise. Further $\boldsymbol{\xi}_i^{(k+1)}$ denotes the $(k+1)$–dimensional vector with elements $\tilde{b}_l$, $l = 0, \ldots, k$ where $\tilde{b}_l = 1$ if $l = i$ and $\tilde{b}_l = 0$ otherwise.

By applying Lemma A.2 we then obtain the elements of the matrix $\mathrm{Cov}(\mathbf{g})$ from

$$\begin{aligned} \mathrm{Cov}(f_i f_j, f_l f_m) &= \mu_i\mu_l\Sigma_{jm} + \mu_i\mu_m\Sigma_{jl} + \mu_j\mu_l\Sigma_{im} + \mu_j\mu_m\Sigma_{il} \\ &\quad + \mu_i\Phi_{jlm} + \mu_j\Phi_{ilm} + \mu_l\Phi_{ijm} + \mu_m\Phi_{ijl} + \Psi_{ijlm} - \Sigma_{ij}\Sigma_{lm} \\ \mathrm{Cov}(f_i f_j, f_l) &= \mu_i\Sigma_{jl} + \mu_j\Sigma_{il} + \Phi_{ijl} \\ \mathrm{Cov}(f_i, f_j) &= \Sigma_{ij} \end{aligned} \tag{3.12}$$

for $i, j, l, m = 1, \ldots, k$ The elements of the vector $\mathrm{Cov}(\mathbf{g}, y)$ are given by

$$\begin{aligned} \mathrm{Cov}(y, f_i f_j) &= \mu_i\Sigma_{0j} + \mu_j\Sigma_{0i} + \Phi_{0ij} \\ \mathrm{Cov}(y, f_i) &= \Sigma_{0i} \end{aligned} \tag{3.13}$$

for $i, j = 1, \ldots, k$ Finally, the necessary expectations are

$$\begin{aligned} \mathrm{E}(f_i f_j) &= \Sigma_{ij} + \mu_i \mu_j \\ \mathrm{E}(f_i) &= \mu_i \\ \mathrm{E}(y) &= \mu_0 \end{aligned} \tag{3.14}$$

for $i, j = 1, \ldots, k$

From the optimal combination parameters the optimal value of the MSPE-function within the class of strong linear plus quadratic combinations can be calculated for given $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ (*goal one* from the introduction). To do so we insert the optimal combination parameters into the general formula for the MSPE-function, which is derived in TROSCHKE and TRENKLER (2000) and given in Appendix B.

Even more important, the regression representation facilitates the application of the strong linear plus quadratic combination to empirical data (*goal two* from the introduction): Here we construct the regression matrix $\mathbf{X}$ from a column of ones (for the constant term), k columns with the squared observations on the individual forecasts, $k(k-1)/2$ columns with the mixed products of the observations and k columns with the observations themselves. The observations on the target variable y yield the vector $\mathbf{y}$. Then we may apply any estimator from linear regression theory to estimate the regression parameters and thus the combination parameters. In our numerical studies for the combination of $k = 2$ forecasts we have observed that using the common least squares estimator $(\widehat{\omega_0}, \widehat{\boldsymbol{\omega}}^{\mathrm{T}})^{\mathrm{T}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$ for this purpose leads to the same results as replacing the true moments $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ by the respective sample moments in the formulae for the optimal combination parameters derived in TROSCHKE and TRENKLER (2000). THIELE (1993, Section 4.2.3) proves a corresponding result for the linear combinations $f_{\mathbf{b},\mathrm{rest}}$ and $f_{\mathbf{b},c,\mathrm{rest}}$.

The advantage of the linear regression representation is that it allows for easier implementation of linear plus quadratic combination for both goals. Now they can be handled for any number k of forecasts to be combined without further effort like in the direct representation developed in TROSCHKE and TRENKLER (2000).

The formulae for the optimal combination parameters in the linear regression representation, however, are even less explicit than in the direct representation where the dependence of $\mathbf{b}_{\mathrm{opt}}$ on $\mathbf{A}_{\mathrm{opt}}$ and of c_{opt} on $\mathbf{b}_{\mathrm{opt}}$ and $\mathbf{A}_{\mathrm{opt}}$ becomes obvious. Thus the regression representation is less suitable for theoretical considerations. Not even the unbiasedness property of the optimal strong linear plus quadratic combination $\mathbf{f}^{\mathrm{T}}\mathbf{A}_{\mathrm{opt}}\mathbf{f} + \mathbf{b}_{\mathrm{opt}}^{\mathrm{T}}\mathbf{f} + c_{\mathrm{opt}}$ could have been concluded from the regression representation. Consequently, the linear regression representation should be used for numerical purposes while the direct representation should be used for theoretical considerations.

3.2 Medium linear plus quadratic combination

The medium linear plus quadratic approach $f_{\mathbf{a},\mathbf{b},c} = \sum_{i=1}^{k} a_i f_i^2 + \mathbf{b}^{\mathrm{T}}\mathbf{f} + c$ emerges from restricting the full matrix $\mathbf{A}$ in the strong approach to a diagonal matrix $\mathrm{dg}(\mathbf{a})$, $\mathbf{a} = (a_1, \ldots, a_k)^{\mathrm{T}} \in \mathbb{R}^k$. Again numerical considerations as well as empirical applications may be facilitated by a regression point of view:
Analogously to the previous section we may regard minimization of the MSPE-function for $f_{\mathbf{a},\mathbf{b},c}$

$$\mathrm{MSPE}(f_{\mathbf{a},\mathbf{b},c}, y) = \mathrm{E}\left[\left(y - \sum_{i=1}^{k} a_i f_i^2 - \sum_{i=1}^{k} b_i f_i - c\right)^2\right] \tag{3.15}$$

as a linear regression problem, namely that of regressing the target variable y on the vector $\mathbf{g} = (g_1, \ldots, g_{2k})^{\mathrm{T}} = (f_1^2, \ldots, f_k^2, f_1, \ldots, f_k)^{\mathrm{T}}$, i.e. on the vector of squared forecasts f_i^2 and forecasts f_i, using a constant term. The coefficients obtained by this regression are the combination parameters: $(\omega_0, \boldsymbol{\omega}^{\mathrm{T}})^{\mathrm{T}} = (\omega_0, \omega_1, \ldots, \omega_{2k})^{\mathrm{T}} = (c, a_1, \ldots, a_k, b_1, \ldots, b_k)^{\mathrm{T}}$.
For the calculation of the theoretically optimal choices $\mathbf{a}_{\mathrm{opt}}$, $\mathbf{b}_{\mathrm{opt}}$ and c_{opt} based on given moments $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ again we use

$$\boldsymbol{\omega} = \mathrm{Cov}(\mathbf{g})^{-1}\,\mathrm{Cov}(\mathbf{g}, y) \quad \text{and} \quad \omega_0 = \mathrm{E}(y) - \boldsymbol{\omega}^{\mathrm{T}}\,\mathrm{E}(\mathbf{g}) \,, \tag{3.16}$$

but of course with a smaller covariance matrix of regression variables $\mathrm{Cov}(\mathbf{g})$ and a smaller vector of covariances between target variable and regression variables $\mathrm{Cov}(\mathbf{g}, y)$ than in the previous case, since the mixed products $f_i f_j$ are excluded here. Equations (3.12), (3.13) and (3.14) are applied again. By inserting the optimal combination parameters into Equation (B.1) we obtain the corresponding optimal MSPE-value within the class of medium linear plus quadratic combinations (*goal one* from the introduction).
For empirical applications (*goal two* from the introduction) we construct the regression matrix $\mathbf{X}$ from a column of ones, k columns with the squared observations on the individual forecasts and k columns with the observations themselves.

3.3 Weak linear plus quadratic combination

In the weak linear plus quadratic approach $f_{\alpha,\mathbf{b},c} = \alpha\mathbf{f}^{\mathrm{T}}\mathbf{f} + \mathbf{b}^{\mathrm{T}}\mathbf{f} + c$ the full matrix $\mathbf{A}$ from the strong approach is restricted to $\alpha\mathbf{I}$, a real scalar multiple of the $k \times k$ identity matrix.
It should be pointed out again, that the weak linear plus quadratic combination increases the number of combination parameters by only one with respect to the

best linear combination, but it involves $k-1$ parameters less than the medium and even $k(k+1)/2-1$ parameters less than the strong linear plus quadratic combination. Consequently, it may be practical in empirical applications where the number of data available for parameter estimation is not large.

Even though it is possible to express the optimal combination parameters and hence also the optimal MSPE-value within the class of weak linear plus quadratic combinations explicitly (cf. TROSCHKE and TRENKLER (2000)), it is reasonable to take a regression point of view for numerical considerations and empirical applications.

Similarly to the previous sections we may regard minimization of the MSPE-function for $f_{\alpha,\mathbf{b},c}$

$$\mathrm{MSPE}(f_{\alpha,\mathbf{b},c}, y) = \mathrm{E}\left[\left(y - \alpha\sum_{i=1}^{k} f_i^2 - \sum_{i=1}^{k} b_i f_i - c\right)^2\right] \tag{3.17}$$

as a linear regression problem, namely that of regressing the target variable y on the vector $\mathbf{g} = (g_1, \ldots, g_{k+1})^{\mathrm{T}} = (\mathbf{f}^{\mathrm{T}}\mathbf{f}, f_1, \ldots, f_k)^{\mathrm{T}}$, i.e. on the vector of the sum of squared forecasts $\mathbf{f}^{\mathrm{T}}\mathbf{f}$ and forecasts f_i, using a constant term. The coefficients obtained by this regression are the combination parameters: $(\omega_0, \boldsymbol{\omega}^{\mathrm{T}})^{\mathrm{T}} = (\omega_0, \omega_1, \ldots, \omega_{k+1})^{\mathrm{T}} = (c, \alpha, b_1, \ldots, b_k)^{\mathrm{T}}$.

For the calculation of the theoretically optimal choices α_{opt}, $\mathbf{b}_{\mathrm{opt}}$ and c_{opt} based on given moments $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ again we use

$$\boldsymbol{\omega} = \mathrm{Cov}(\mathbf{g})^{-1}\,\mathrm{Cov}(\mathbf{g}, y) \quad \text{and} \quad \omega_0 = \mathrm{E}(y) - \boldsymbol{\omega}^{\mathrm{T}}\,\mathrm{E}(\mathbf{g}), \tag{3.18}$$

but of course with an even smaller covariance matrix of regression variables $\mathrm{Cov}(\mathbf{g})$ and a smaller vector of covariances between target variable and regression variables $\mathrm{Cov}(\mathbf{g}, y)$ than before. Equations (3.12), (3.13) and (3.14) are applied again after using the bilinearity of the covariance operator and the linearity of the expectation operator

$$\begin{aligned}
\mathrm{Cov}(\mathbf{f}^{\mathrm{T}}\mathbf{f}, y) &= \sum_{i=1}^{k} \mathrm{Cov}(f_i^2, y) \\
\mathrm{Cov}(\mathbf{f}^{\mathrm{T}}\mathbf{f}, f_j) &= \sum_{i=1}^{k} \mathrm{Cov}(f_i^2, f_j) \\
\mathrm{Cov}(\mathbf{f}^{\mathrm{T}}\mathbf{f}, \mathbf{f}^{\mathrm{T}}\mathbf{f}) &= \sum_{i=1}^{k}\sum_{j=1}^{k} \mathrm{Cov}(f_i^2, f_j^2) \\
\mathrm{E}(\mathbf{f}^{\mathrm{T}}\mathbf{f}) &= \sum_{i=1}^{k} \mathrm{E}(f_i^2)\,.
\end{aligned} \tag{3.19}$$

Inserting the optimal combination parameters into Equation (B.1) leads to the corresponding optimal MSPE-value within the class of weak linear plus quadratic combinations (*goal one* from the introduction).
For empirical applications (*goal two* from the introduction) we build the regression matrix $\mathbf{X}$ from a column of ones, a column with the sum of the squared observations on the individual forecasts and k columns with the observations themselves.

Since additional explanatory variables have been included in the regression, all three versions of optimal linear plus quadratic combination are superior to all linear combinations in theory and, consequently, have the potential to outperform them in empirical applications.

3.4 Linear plus quadratic adjustment

All three linear plus quadratic combined forecasts coincide in the case of $k = 1$ forecast, i.e. we only need to consider one *linear plus quadratic adjustment*

$$(f_i)_{\alpha,b,c} = \alpha f_i^2 + b f_i + c \tag{3.20}$$

with $\alpha, b, c \in \mathbb{R}$. The corresponding regression models are obvious from the preceding subsections.

In Section 4 we will report about first investigations on the quality of all the above linear and linear plus quadratic adjustments and combinations in the case of $k = 2$ forecasts.

4 Empirical and theoretical comparisons

In this section we will present an empirical example illustrating the various adjustments of single forecasts as well as the combination of $k = 2$ forecasts on the basis of the new methods. This will be followed by a theoretical comparison of these methods based on a given set of moments $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ obtained from the data of the example. It should be pointed out, however, that these comparisons are only meant to provide a first impression of the possible usefulness of the linear plus quadratic approaches. Detailed analyses are bound to follow, and they will be presented in a future paper.
The data for the numerical example are taken from a larger data set of German macro economic variables and corresponding forecasts investigated by Klapper

Year	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986
y	3.6	3.1	3.4	3.2	1.7	-1.2	-2.2	1.1	0.6	1.8	4.3
f_1	3.0	4.5	3	3.5	2.0	1	-0.5	-0.5	0	1.5	3.5
f_2	2.5	4.5	3	3.5	1.5	1	0.0	-0.5	1	1.5	3.0
Year	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	
y	3.5	2.7	1.7	4.7	3.6	1.7	0.2	0.6	1.8	1.3	
f_1	3.0	3.0	2.0	3.5	3.5	2	0	-1.5	0.5	2.0	
f_2	3.5	2.5	2.5	4.0	3.0	2	0	-1.0	0.5	2.5	

Table 1: Real change of German private consumption (y) and corresponding DIW and Ifo forecasts (f_1, f_2) for the period from 1976 to 1996

(1998). We picked out the DIW (Deutsches Institut für Wirtschaftsforschung, f_1) and Ifo (Ifo-Institut für Wirtschaftsforschung, f_2) forecasts for the target variable 'real change of private consumption' (y). These yearly data are available for a period of 21 years from 1976 to 1996. They are given in Table 1.

When evaluating the data it is important to take their availability into account: The forecasts f_1 and f_2 for year t, say, are made at the end of year $t-1$ and the true value of the target variable y for the year $t-2$ are not published by the Statistisches Bundesamt before the end of year $t-1$. Consequently, at the time when the individual forecasts for year t are to be combined, namely at the end of year $t-1$, we can only use the past data up to year $t-2$.

These past data serve to estimate the optimal combination parameters at each point of time with the help of the regression models for empirical data from the previous sections. Due to structural changes in the data set the optimal combination parameters may not be stable over time. A common procedure in this situation is to use only the latest observations for parameter estimation. Of course the amount of past data should not be too small either so that the regression fit is at least fairly reasonable. As a compromise we chose a history of 10 data points for parameter estimation.

Altogether we will use the data from 1976 to 1985 to estimate the combination parameters for the 1987 forecasts, the data from 1977 to 1986 to estimate the combination parameters for the 1988 forecasts, and so on. This leads to a time span of 10 years (1987 to 1996) in which the performance of the various methods is evaluated by means of the average of the squared forecast errors. This is the empirical counterpart of mean square prediction error and will consequently be denoted as $\widehat{\text{MSPE}}$.

A very simple strategy for the combination of the single forecasts is their arithmetic

Forecast $f.$		$\widehat{\text{MSPE}}(f., y)$
DIW forecast	$f_1 = f_{\text{DIW}}$	1.14
Adjustments:	$f_{\text{DIW},\hat{\alpha}_{\text{opt}},\hat{b}_{\text{opt}},\hat{c}_{\text{opt}}}$	0.61
	$f_{\text{DIW},\hat{b}_{\text{opt}},\hat{c}_{\text{opt}}}$	0.83
	$f_{\text{DIW},\hat{b}_{\text{opt}}}$	1.30
	$f_{\text{DIW},1,\hat{c}_{\text{opt}}}$	1.01
Ifo forecast	$f_2 = f_{\text{Ifo}}$	0.97
Adjustments:	$f_{\text{Ifo},\hat{\alpha}_{\text{opt}},\hat{b}_{\text{opt}},\hat{c}_{\text{opt}}}$	0.60
	$f_{\text{Ifo},\hat{b}_{\text{opt}},\hat{c}_{\text{opt}}}$	0.93
	$f_{\text{Ifo},\hat{b}_{\text{opt}}}$	1.11
	$f_{\text{Ifo},1,\hat{c}_{\text{opt}}}$	0.99
Linear combinations:	$f_{\hat{\mathbf{b}}_{\text{opt}},\hat{c}_{\text{opt}}}$	1.03
	$f_{\hat{\mathbf{b}}_{\text{opt}}}$	1.41
	$f_{\hat{\mathbf{b}}_{\text{opt}},\text{rest}}$	1.16
	$f_{\hat{\mathbf{b}}_{\text{opt}},\hat{c}_{\text{opt}},\text{rest}}$	1.10
LPQ combinations:	$f_{\hat{\mathbf{A}}_{\text{opt}},\hat{\mathbf{b}}_{\text{opt}},\hat{c}_{\text{opt}}}$	1.14
	$f_{\hat{\mathbf{a}}_{\text{opt}},\hat{\mathbf{b}}_{\text{opt}},\hat{c}_{\text{opt}}}$	0.66
	$f_{\hat{\alpha}_{\text{opt}},\hat{\mathbf{b}}_{\text{opt}},\hat{c}_{\text{opt}}}$	0.64

Table 2: $\widehat{\text{MSPE}}$-values of adjusted and combined forecasts in an empirical application (all values relative to the $\widehat{\text{MSPE}}$ of the arithmetic mean)

mean. Since it is easy to apply and also quite successful in empirical investigations, any other combination technique is measured against the arithmetic mean. Therefore we decided to present all $\widehat{\text{MSPE}}$-values relative to the $\widehat{\text{MSPE}}$-value of the arithmetic mean, which is 0.7538 in the considered time period. All decimals have been deleted following the second decimal such that methods outperforming the arithmetic mean can be identified immediately. Proceeding in this way makes the results directly comparable to those in Klapper (1998).

The results of this evaluation are presented in Table 2. It can be seen that in this example the weak linear plus quadratic combination is the best of all combination methods followed by the medium linear plus quadratic combination technique. Only these two combinations perform better than both individual forecasts. Their $\widehat{\text{MSPE}}$-values are about two third of the value for the arithmetic mean. The best linear combination technique is the unrestricted combination with constant term which is about as good as the arithmetic mean. The strong linear plus quadratic combination performs equally to the worse of the two individual forecasts which is 14% worse than the arithmetic mean. The commonly used linear combination without constant term and with restriction on the elements of $\mathbf{b}$ performs even slightly worse.

The best adjustments of individual forecasts in this example are the linear plus quadratic adjustments. Both of them have $\widehat{\text{MSPE}}$-values of about 60% of the value belonging to the arithmetic mean. Thus they are the best of the considered techniques even better than all combination methods. Also the linear unrestricted adjustments with constant term perform quite well.

It is interesting to note that the linear unrestricted adjustments and combination without constant term perform worst in their respective groups.

The forecasts for the years 1987 to 1996 produced by the weak linear plus quadratic combination of f_1 and f_2 are given by 2.4075, 2.9264, 1.6082, 4.2094, 4.1306, 1.4047, 0.0789, 1.6358, 0.5785 and 1.9407. Together with the target variable, the individual forecasts and their arithmetic mean they are visualized in Figure 1.

It should be pointed out that the preceding analysis represents only a single example and cannot be generalized. In our first investigations there have been examples where the linear plus quadratic techniques, especially the strong linear plus quadratic technique, perform significantly worse. Presumably this is due to the very small amount of past data available for the regression, only 10 data points seem to be very little. Again we must refer to a more detailed analysis of the performance of the linear plus quadratic techniques which is bound to follow.

To judge the potential of the linear plus quadratic techniques it is interesting to compare the optimal MSPE-values within the various approaches for the case where the moments $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ of the joint distribution of y, f_1 and f_2 are known. In order to base these considerations on realistic grounds we are now going to use the sample moments, which may be calculated from the whole set of 21 data points in Table 1, as the true moments.

From these moments we may then determine the optimal adjustment or combination parameters belonging to the different methods. Following the formulae from Section 3.1 we obtain, for example, the optimal parameters for the strong linear plus quadratic combination approach:

$$\mathbf{A}_{\text{opt}} = \begin{pmatrix} 2.3910 & -2.7544 \\ -2.7544 & 3.3331 \end{pmatrix}, \quad \mathbf{b}_{\text{opt}} = \begin{pmatrix} 3.3049 \\ -3.3753 \end{pmatrix}, \quad \mathbf{c}_{\text{opt}} = 0\ .6113. \tag{4.1}$$

Inserting the respective optimal combination parameters into the general MSPE-function (B.1) we derive the optimal MSPE-values for all the considered methods.

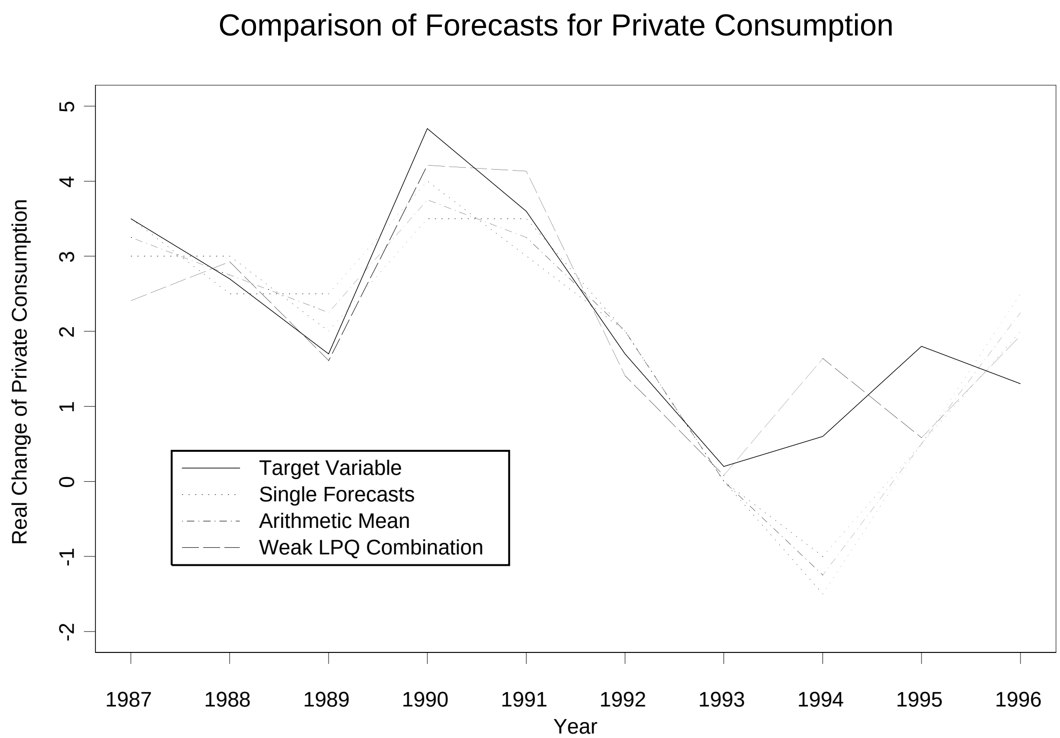

Figure 1: Target variable *real change of private consumption*, together with its DIW and Ifo forecasts, their arithmetic mean and their weak linear plus quadratic combination.

Forecast $f_\cdot$		$\mathrm{MSPE}(f_\cdot, y)$
DIW forecast	$f_1 = f_{\mathrm{DIW}}$	0.98
Adjustments:	$f_{\mathrm{DIW},\alpha_{\mathrm{opt}},b_{\mathrm{opt}},c_{\mathrm{opt}}}$	0.88
	$f_{\mathrm{DIW},b_{\mathrm{opt}},c_{\mathrm{opt}}}$	0.93
	$f_{\mathrm{DIW},b_{\mathrm{opt}}}$	0.98
	$f_{\mathrm{DIW},1,c_{\mathrm{opt}}}$	0.97
Ifo forecast	$f_2 = f_{\mathrm{Ifo}}$	1.09
Adjustments:	$f_{\mathrm{Ifo},\alpha_{\mathrm{opt}},b_{\mathrm{opt}},c_{\mathrm{opt}}}$	1.04
	$f_{\mathrm{Ifo},b_{\mathrm{opt}},c_{\mathrm{opt}}}$	1.06
	$f_{\mathrm{Ifo},b_{\mathrm{opt}}}$	1.08
	$f_{\mathrm{Ifo},1,c_{\mathrm{opt}}}$	1.09
Linear combinations:	$f_{\mathbf{b}_{\mathrm{opt}},c_{\mathrm{opt}}}$	0.92
	$f_{\mathbf{b}_{\mathrm{opt}}}$	0.98
	$f_{\mathbf{b}_{\mathrm{opt}},\mathrm{rest}}$	0.98
	$f_{\mathbf{b}_{\mathrm{opt}},c_{\mathrm{opt}},\mathrm{rest}}$	0.97
LPQ combinations:	$f_{\mathbf{A}_{\mathrm{opt}},\mathbf{b}_{\mathrm{opt}},c_{\mathrm{opt}}}$	0.73
	$f_{\mathbf{a}_{\mathrm{opt}},\mathbf{b}_{\mathrm{opt}},c_{\mathrm{opt}}}$	0.86
	$f_{\alpha_{\mathrm{opt}},\mathbf{b}_{\mathrm{opt}},c_{\mathrm{opt}}}$	0.86

Table 3: MSPE-values of adjusted and combined forecasts for certain known moments $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ (all values relative to the MSPE of the arithmetic mean)

Note that a linear combination is a linear plus quadratic combination with $\mathbf{A} = \mathbf{0}$. For all adjustments and all combinations except strong and medium linear plus quadratic combination we might as well use the respective direct formulae for the optimal MSPE-value developed in TROSCHKE and TRENKLER (2000) for that purpose.

Again we report all these MSPE-values relative to the MSPE of the arithmetic mean, which is 1.0894. All values in Table 3 have been deleted after the second decimal.

Since the moments $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are assumed to be known, the calculations can be done on a theoretical basis and, hence, the MSPE-values reflect the theoretical ranking of the various methods: strong linear plus quadratic combination is not worse than medium linear plus quadratic combination, which in turn is not worse than weak linear plus quadratic combination, which in turn is not worse than the linear unrestricted combination with constant term, and so on.

In the situation under consideration the expected squared error loss of the strong linear plus quadratic combination is 27% less than that of the arithmetic mean.

Medium and weak linear plus quadratic combinations are expected to be only 14% better than the arithmetic mean. We may conclude that in the above application medium and weak linear plus quadratic combinations performed much better than might have been expected, especially when taking into consideration that the necessity to estimate the optimal combination parameters leads to an even worse theoretical MSPE. In addition the linear plus quadratic adjustments performed much better in the application than might have been expected.
It can be seen that there is some potential in the linear plus quadratic approaches to outperform the arithmetic mean. How well this potential is exploited will depend on how good the regression reflects the true relationship between target variable y and forecasts f_i. Clearly, the more suitable data are available for that regression, the better. Consequently, the linear plus quadratic approaches should be more valuable for monthly, weekly or even daily data (e.g. from the stock market) than they are for yearly data. Also the data should not be subject to extreme structural changes during the considered period.

5 Conclusions

In this paper we have introduced the linear regression approach for the linear plus quadratic combination of forecasts. We have also considered the classical linear approaches as competitors to the new approaches as well as adjustments of individual forecasts which emerge from the special case $k = 1$. The most important advantage of the regression approach is that it allows for easy implementation for any number k of forecasts to be combined.
Furthermore, we have reported on first comparisons of the classical and the new approaches in a small example. A realistic empirical situation was considered on the one hand. On the other hand a numerical comparison of the optimal MSPE-values possible based on given moments of the joint distribution of $\mathbf{y}$ and $\mathbf{f}$ was carried out. For the latter each of the linear plus quadratic approaches requires knowledge about the moments up to order four, whereas linear approaches only need the moments up to order two to be known.
We have seen that employing linear plus quadratic adjustments and combinations may be beneficial, but also that this is not always the case. Due to the smaller number of parameters involved the weak linear plus quadratic combination seems to be suitable if only a small amount of data is available for combination parameter estimation.
A much more detailed analysis of the possible benefits of the linear plus quadratic

approaches has to follow, as was explained in Section 4. It will be carried out in a follow-up paper by the same authors. A point of special interest would be to find a guideline for potential users identifying situations beforehand in which linear plus quadratic combination of forecasts is promising. Especially the question of how much data should be available is interesting. Another point is to find out whether it is worthwhile to consider the combination of more than $k = 2$ forecasts via the linear plus quadratic approaches. It may also be interesting to generalize the linear plus quadratic approaches to the combination of multivariate forecasts, i.e. to the situation where each forecaster does not only predict the outcome of one variable but of a set of variables.

Appendix

A Two useful results

This section lists two results which are important for our considerations. The first lemma gives explicit representations of some matrix or vector expressions in terms of the elements involved.

Lemma A.1 *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*

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

The second lemma is concerned with the first and second order moments of quadratic forms. It should be pointed out that no distributional assumption is made. Assuming (multivariate) normality would lead to much simpler formulae on the one hand. But on the other hand the normality assumption would render the whole linear plus quadratic approach to the combination of forecasts unnecessary, as has been made clear in the introduction.

Lemma A.2 (Rao and Kleffe, 1988, p. 32, (iv)) *Let* $\tilde{\mathbf{Y}} = \tilde{\boldsymbol{\mu}} + \tilde{\boldsymbol{\varepsilon}}$ *where* $\tilde{\boldsymbol{\mu}}$ *is a constant vector and* $\tilde{\boldsymbol{\varepsilon}}$ *is a vector random variable with moments* $\mathrm{E}(\tilde{\boldsymbol{\varepsilon}}) = \mathbf{0}$, $\mathrm{E}(\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^{\mathrm{T}}) = \tilde{\boldsymbol{\Sigma}}$, $\mathrm{E}(\tilde{\boldsymbol{\varepsilon}} \otimes \tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^{\mathrm{T}}) = \tilde{\boldsymbol{\Phi}}$ *and* $\mathrm{E}(\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^{\mathrm{T}} \otimes \tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^{\mathrm{T}}) = \tilde{\boldsymbol{\Psi}}$*. Further let* $\tilde{\mathbf{a}}$ *and* $\tilde{\mathbf{b}}$ *be vectors and let* $\tilde{\mathbf{A}}$ *and* $\tilde{\mathbf{B}}$ *be symmetric matrices of appropriate dimensions. Then*

(a) $\mathrm{E}(\tilde{\mathbf{a}}^{\mathrm{T}}\tilde{\mathbf{Y}} + \tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{A}}\tilde{\mathbf{Y}}) = \tilde{\mathbf{a}}^{\mathrm{T}}\tilde{\boldsymbol{\mu}} + \tilde{\boldsymbol{\mu}}^{\mathrm{T}}\tilde{\mathbf{A}}\tilde{\boldsymbol{\mu}} + \mathrm{tr}(\tilde{\mathbf{A}}\tilde{\boldsymbol{\Sigma}})$,

(b) $\mathrm{Cov}(\tilde{\mathbf{a}}^{\mathrm{T}}\tilde{\mathbf{Y}} + \tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{A}}\tilde{\mathbf{Y}}, \tilde{\mathbf{b}}^{\mathrm{T}}\tilde{\mathbf{Y}} + \tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{B}}\tilde{\mathbf{Y}})$

$$\begin{aligned}
&= \tilde{\mathbf{b}}^{\mathrm{T}}\left[2\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{A}}\tilde{\boldsymbol{\mu}} + \tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{a}} + \tilde{\boldsymbol{\Phi}}^{*}(\tilde{\mathbf{A}})\right] \\
&+ \mathrm{tr}\Big(\tilde{\mathbf{B}}\Big[4\tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\tilde{\mathbf{A}}\tilde{\boldsymbol{\Sigma}} + 2\tilde{\boldsymbol{\Phi}}(\tilde{\mathbf{A}}\tilde{\boldsymbol{\mu}}) + 2\tilde{\boldsymbol{\Phi}}^{*}(\tilde{\mathbf{A}})\tilde{\boldsymbol{\mu}}^{\mathrm{T}} \\
&\qquad + \tilde{\boldsymbol{\Psi}}(\tilde{\mathbf{A}}) + 2\tilde{\boldsymbol{\mu}}\tilde{\mathbf{a}}^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}} + \tilde{\boldsymbol{\Phi}}(\tilde{\mathbf{a}}) - \mathrm{tr}(\tilde{\mathbf{A}}\tilde{\boldsymbol{\Sigma}})\tilde{\boldsymbol{\Sigma}}\Big]\Big) \ .
\end{aligned}$$

Here the following abbreviations have been used: For a vector $\tilde{\mathbf{c}} = (\tilde{c}_i)$ *and a matrix* $\tilde{\mathbf{C}} = (\tilde{c}_{ij})$ *we define*

$$
\begin{aligned}
\tilde{\boldsymbol{\Psi}}(\tilde{\mathbf{C}}) &= \sum_i \sum_j \tilde{c}_{ij} \tilde{\boldsymbol{\Psi}}_{ij} , \\
\tilde{\boldsymbol{\Phi}}(\tilde{\mathbf{c}}) &= \sum_i \tilde{c}_i \tilde{\boldsymbol{\Phi}}_i , \\
\tilde{\boldsymbol{\Phi}}^*(\tilde{\mathbf{C}}) &= (\mathrm{tr}(\tilde{\mathbf{C}}\tilde{\boldsymbol{\Phi}}_i))_i ,
\end{aligned}
$$

i.e. the first two quantities are matrices, whereas the last one is a vector.

B General MSPE-function for linear plus quadratic combinations

In Troschke and Trenkler (2000) the authors derive the following expression for the mean square prediction error of a general linear plus quadratic combination $f_{\mathbf{A},\mathbf{b},c}$, where the terms have been ordered with respect to the occurring unknowns:

$$
\begin{aligned}
&\mathrm{MSPE}(f_{\mathbf{A},\mathbf{b},c}, y) = \\
&= 4\,\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{ff}}\mathbf{A}\boldsymbol{\mu}_{\mathbf{f}} + 4\,\boldsymbol{\varphi}_{\mathbf{A}}^{\mathrm{T}}\mathbf{A}\boldsymbol{\mu}_{\mathbf{f}} + \mathrm{tr}(\mathbf{A}\boldsymbol{\psi}_{\mathbf{A}}) + (\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}\mathbf{A}\boldsymbol{\mu}_{\mathbf{f}})^2 + 2\,\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}\mathbf{A}\boldsymbol{\mu}_{\mathbf{f}}\,\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{ff}}) \\
&\quad - 4\boldsymbol{\Sigma}_{\mathbf{f}0}^{\mathrm{T}}\mathbf{A}\boldsymbol{\mu}_{\mathbf{f}} - 2\,\mathrm{tr}(\mathbf{A}\boldsymbol{\Phi}_{0\mathbf{ff}}) - 2\mu_0\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}\mathbf{A}\boldsymbol{\mu}_{\mathbf{f}} - 2\mu_0\,\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{ff}}) \\
&\quad + 4\,\mathbf{b}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{ff}}\mathbf{A}\boldsymbol{\mu}_{\mathbf{f}} + 2\,\mathbf{b}^{\mathrm{T}}\boldsymbol{\varphi}_{\mathbf{A}} + 2\,\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}\mathbf{A}\boldsymbol{\mu}_{\mathbf{f}}\mathbf{b}^{\mathrm{T}}\boldsymbol{\mu}_{\mathbf{f}} + 2\,\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{ff}})\mathbf{b}^{\mathrm{T}}\boldsymbol{\mu}_{\mathbf{f}} \\
&\quad + \mathbf{b}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{ff}}\mathbf{b} + \mathbf{b}^{\mathrm{T}}\boldsymbol{\mu}_{\mathbf{f}}\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}\mathbf{b} \\
&\quad - 2\mathbf{b}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{f}0} - 2\mu_0\mathbf{b}^{\mathrm{T}}\boldsymbol{\mu}_{\mathbf{f}} \\
&\quad + 2\,\boldsymbol{\mu}_{\mathbf{f}}^{\mathrm{T}}\mathbf{A}\boldsymbol{\mu}_{\mathbf{f}}c + 2\,\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{ff}})c \\
&\quad + 2\,\mathbf{b}^{\mathrm{T}}\boldsymbol{\mu}_{\mathbf{f}}c \\
&\quad + c^2 \\
&\quad - 2\mu_0 c \\
&\quad + \Sigma_{00} + \mu_0^2 ,
\end{aligned}
\tag{B.1}
$$

where

$$
\boldsymbol{\varphi}_{\mathbf{A}} = \begin{pmatrix} \mathrm{tr}(\mathbf{A}\boldsymbol{\Phi}_{1\mathbf{ff}}) \\ \vdots \\ \mathrm{tr}(\mathbf{A}\boldsymbol{\Phi}_{k\mathbf{ff}}) \end{pmatrix}
\tag{B.2}
$$

is a k–dimensional vector and

$$\boldsymbol{\psi}_{\mathbf{A}} = \begin{pmatrix} \mathrm{tr}(\mathbf{A}\boldsymbol{\Psi}_{11\mathbf{ff}}) & \dots & \mathrm{tr}(\mathbf{A}\boldsymbol{\Psi}_{1k\mathbf{ff}}) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(\mathbf{A}\boldsymbol{\Psi}_{k1\mathbf{ff}}) & \dots & \mathrm{tr}(\mathbf{A}\boldsymbol{\Psi}_{kk\mathbf{ff}}) \end{pmatrix} \tag{B.3}$$

is a symmetric $k \times k$ matrix.

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

References

BATES, J.M. and GRANGER, C.W.J. (1969): 'The combination of forecasts'. *Operational Research Quarterly* **20**, 451-468.

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

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.

KLAPPER, M. (1998): Combining German Macro Economic Forecasts Using Rank-Based Techniques. *Technical Report 19/1998, Sonderforschungsbereich 475, University of Dortmund.*

RAO, C.R. (1965): Linear Statistical Inference and Its Applications. *Wiley, New York.*

RAO, C.R. and KLEFFE, J. (1988): Estimation of Variance Components and Applications. *North-Holland, Amsterdam.*

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

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.*