A Simulation Study To Compare Various Covariance Adjustment Techniques

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

Abstract: A common procedure when combining two multivariate unbiased estimates (or forecasts) is the covariance adjustment technique (CAT). Here the optimal combination weights depend on the covariance structure of the estimators. In practical applications, however, this covariance structure is hardly ever known and, thus, has to be estimated. An effect of this drawback may be that the theoretically best method is no longer the best. In a simulation study (using normally distributed data) three different variants of CAT are compared with respect to their accuracy. These variants are different in the portion of the covariance structure that is estimated. We characterize which variant is appropriate in different situations and quantify the gains and losses that occur.

Keywords: Combination of multivariate estimators, covariance adjustment technique.

AMS 1991 Subject Classification: 62H12, 62F10

1 Introduction

Let $\mathbf{T}_1$ and $\mathbf{T}_2$ be two unbiased estimators of a parameter vector $\boldsymbol{\theta} \in I\!R^k$, i.e.

$$\mathrm{E}(\mathbf{T}_1) = \boldsymbol{\theta} = \mathrm{E}(\mathbf{T}_2) \ ,$$

with covariance structure

$$\mathrm{Cov}\begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^\top & \boldsymbol{\Sigma}_{22} \end{pmatrix} =: \boldsymbol{\Sigma} \ .$$

A problem frequently arising in statistics is to combine these two estimators in order to obtain a better estimator of $\boldsymbol{\theta}$. The idea behind combining is that each of the two estimators uses some information on $\boldsymbol{\theta}$ that the other neglects.

A common procedure to combine two multivariate unbiased estimators is the so called covariance adjustment technique (CAT). In this report we will investigate the performance of three variants of this technique in the case where the covariance matrix $\mathbf{\Sigma}$ is unknown and, thus, has to be estimated.

Section 2 will introduce the three variants of CAT considered here. Sections 3 and 4 describe a simulation study carried out to compare these techniques. Random data from a 4-variate normal distribution will be used here. Finally, Section 5 gives some concluding remarks.

In a further technical report (Troschke (1999)) the application of covariance adjustment techniques to empirical data, namely to German macro economic forecast data is investigated, showing that CAT is also applicable for predictions.

2 Covariance Adjustment Methods

Let $\mathbf{T}_1$ and $\mathbf{T}_2$ be two unbiased estimators of a parameter vector $\boldsymbol{\theta} \in I\!R^k$. The common point in all three covariance adjustment methods described below is that we are trying to find the optimal combination $\mathbf{T}_c$ of $\mathbf{T}_1$ and $\mathbf{T}_2$ in the sense of a linear combination

$$\mathbf{T}_c = \mathbf{L}_1\mathbf{T}_1 + \mathbf{L}_2\mathbf{T}_2 \ ,$$

where $\mathbf{L}_1, \mathbf{L}_2$ are $k \times k$ real matrices.

In order to make the combined estimator unbiased, the combination weights must add to unity, i.e. $\mathbf{L}_1 + \mathbf{L}_2 = \mathbf{I}$, where $\mathbf{I}$ is the $k \times k$ identity matrix. This means that we are looking for an optimal combination of the type

$$\mathbf{T}_\mathbf{L} = (\mathbf{I} - \mathbf{L})\mathbf{T}_1 + \mathbf{L}\mathbf{T}_2 \ ,$$

with $\mathbf{L} \in I\!R^{k\times k}$.

Following the concept of Rao (1966, 1967) $\mathbf{L}$ has to be chosen from $I\!R^{k\times k}$, the set of *all* $k \times k$ real matrices, such that the covariance matrix of the combined estimator $\mathbf{T}_\mathbf{L}$ is minimal with respect to the Löwner ordering. (For $\mathbf{A}, \mathbf{B} \in I\!R^{k\times k}$ we call $\mathbf{A}$ lower than or equal to $\mathbf{B}$ with respect to the Löwner ordering if $\mathbf{B} - \mathbf{A}$ is nonnegative definite, cf. Löwner (1934).) In this setting the optimal choice for $\mathbf{L}$ is

$$\mathbf{L}_0 = (\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12})(\mathbf{\Sigma}_{11} + \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12} - \mathbf{\Sigma}_{12}^\top)^{-1} \ .$$

Rao (1966) coined the notion *Covariance Adjustment* for this procedure. We are going to refer to it as the *Strong Covariance Adjustment Technique (SCAT)* in the following.

The above method requires assignment of the full weight matrix $\mathbf{L} \in I\!R^{k\times k}$, i.e. determining k^2 parameters. The idea of Trenkler and Ihorst (1995) was to reduce this number substantially and, thus, to provide a more feasible combination procedure. They restricted $\mathbf{L}$ to be *a multiple of the identity matrix*, i.e. $\mathbf{L} = \alpha\mathbf{I}$, $\alpha \in I\!R$. Consequently we are looking for an optimal choice of α in

$$\mathbf{T}_\alpha = (1-\alpha)\mathbf{T}_1 + \alpha\mathbf{T}_2 \ .$$

Restricting $\mathbf{L}$ in such a way, minimization with respect to the covariance matrix causes difficulties, cf. Odell et al., p. 1632, (1989). Hence, Trenkler and Ihorst (1995) chose the total variance as a minimization criterion. Thus, a scalar optimization criterion has been selected for the scalar linearly combined model. The optimal choice of α with respect to the total variance is

$$\alpha_0 = \mathrm{tr}(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12})(\mathrm{tr}(\mathbf{\Sigma}_{11} + \mathbf{\Sigma}_{22} - 2\mathbf{\Sigma}_{12}))^{-1} \ .$$

Obviously, this method requires assignment of only one parameter. It will be referred to as the *weak covariance adjustment technique (WCAT)*.

An intermediate method between the two extremes would be to restrict $\mathbf{L}$ to the set of diagonal matrices, i.e. to look for a $k \times k$ diagonal matrix $\mathbf{D}$ minimizing the covariance matrix of

$$\mathbf{T}_\mathbf{D} = (\mathbf{I} - \mathbf{D})\mathbf{T}_1 + \mathbf{D}\mathbf{T}_2$$

with respect to the Löwner ordering. The choice of diagonal matrix weights effects that each component of the forecasts is regarded separately. Hence, the univariate problem has to be solved for each of the k components. Consequently, the diagonal of the optimal matrix weight $\mathbf{D}_0$ consists of the respective optimal univariate choices, i.e.

$$\mathbf{D}_0 = \mathrm{diag}(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12})(\mathrm{diag}(\mathbf{\Sigma}_{11} + \mathbf{\Sigma}_{22} - 2\mathbf{\Sigma}_{12}))^{-1} \ ,$$

(compare e.g. Bates and Granger (1969)). Here k parameters have to be assigned and, therefore, we will refer to this method as *medium covariance adjustment technique (MCAT)*.

Other variants of covariance adjustment could be thought of including estimation of another portion of the covariance matrix, but the three methods considered here are the most obvious.

Clearly, according to the sets from which the matrix weights are chosen, SCAT has the best theoretical properties, followed by MCAT and then WCAT. Note that if one estimator is better than another with respect to the covariance matrix criterion then it is also better with respect to total variance criterion. Note further that the covariance matrix criterion coincides with the matrix mean square error criterion (MMSE) with respect to $\boldsymbol{\theta}$ and the total variance criterion coincides with the scalar mean square error criterion (SMSE) with respect to $\boldsymbol{\theta}$. The reason for this is that $\mathbf{T}_1$ and $\mathbf{T}_2$ are unbiased and hence, by the choice of the combination weights, also $\mathbf{T}_{\mathbf{L}}, \mathbf{T}_{\mathbf{D}}$ and $\mathbf{T}_\alpha$ are unbiased for $\boldsymbol{\theta}$.

In most practical applications the covariance structure $\boldsymbol{\Sigma}$ is not known, and therefore the optimal weights for the covariance adjustment techniques are not known either. Consequently, the optimal combination weights α_0, $\mathbf{D}_0$ and $\mathbf{L}_0$ have to be estimated. Hence, the ranking of the three procedures might change in empirical applications, especially because the number of parameters linked to these weights is different $(1, k, k^2)$.

Reasoning whether one should rather use the strong or the weak CAT when $\boldsymbol{\Sigma}$ is not known, Trenkler and Ihorst (1995, pp. 191–192) state: *In such a case it seems advantageous to apply the weak covariance adjustment technique, because instead of estimating a matrix we need only estimate the scalar α. The number of parameters in $\widehat{\boldsymbol{\Sigma}}$ used for $\mathbf{T}_\alpha$ is thereby reduced substantially.*

Keeping in mind that the overall best choice for the combination weight is $\mathbf{L}_0$, there are also arguments against WCAT: Even if the estimator $\widehat{\mathbf{L}}_0$ for the SCAT optimal weight $\mathbf{L}_0$ is bad, that does not necessarily mean that $\widehat{\alpha}_0\mathbf{I}$ is a better estimator, since $\mathbf{L}_0$ need not be near to a diagonal matrix.

However, we can argue in favor of WCAT that neither of the estimated weights will be optimal, but the effort for estimating the WCAT optimal α is much smaller. Still, with WCAT we may hope for an improvement over the single estimators or their arithmetic mean. Furthermore, there may be situations where the optimal $\mathbf{L}_0$ is close to a multiple of the identity matrix $\mathbf{I}$, which might favor the use of WCAT. If, e.g., $\boldsymbol{\Sigma}_{12} \approx \mathbf{0}$, $\boldsymbol{\Sigma}_{11} \approx \sigma_1^2\mathbf{I}$ and $\boldsymbol{\Sigma}_{22} \approx \sigma_2^2\mathbf{I}$ then we have this situation with $\mathbf{L}_0 \approx \sigma_1^2(\sigma_1^2 + \sigma_2^2)^{-1}\mathbf{I}$.

Another reason why SCAT may not be so successful as could be hoped for is the following: In the literature on combination of univariate forecasts it is a frequently stated observation, that estimating the optimal combination weights neglecting covariances between the estimators often leads to better combined estimates than calculation employing these covariances (see e.g. Makridakis and Winkler (1983)). Therefore, it may well be the case that employing covariances between the different components of the estimators is not beneficial as well. These covariances are employed by SCAT but not by MCAT or WCAT.

At first glance it may seem counterintuitive that other estimators could produce better results in practice than SCAT which uses the full covariance structure. If the true optimal combination weight is unknown and consequently has to be estimated, however, the quality of the weight estimator plays an important role: This estimator may or may not be unbiased and it may have a large or small variance. For the weak covariance adjustment technique Trenkler and Ihorst (1995) show what amount of accuracy is gained by the WCAT procedure and what portion of it is lost again by the necessity to estimate the optimal combination parameter. If the SCAT optimal combination weight cannot be estimated satisfactorily, other techniques like MCAT or WCAT may work better, or it may be that weight estimation without covariances has better properties than estimation employing covariances between the parameter estimators.

The performance of the three covariance adjustment methods shall now be investigated, when $\mathbf{\Sigma}$ has to be estimated. In order to do so a simulation study has been carried out and is described in the following two sections. To the best knowledge of the author this is the first numerical comparison of these methods. Not only will it be interesting to see which variant performs best in various situations, but also *how much* may be gained by using one technique instead of another. A comparison of the covariance adjustment techniques using empirical data will be described in Troschke (1999).

3 Simulation Study

The goal of the following study is to find out whether SCAT, which is the theoretically best variant of CAT, is the best variant in practice. If so, it is interesting by what margin SCAT outperforms the other variants. On the other hand we might find evidence for the conjecture that WCAT and MCAT perform better, since the number of parameters to be estimated is substantially smaller. Other interesting issues will also be addressed, like the question whether it is profitable to employ the covariance between the individual estimators. Presumably, the results will depend heavily on the chosen covariance structure and sample size.

This simulation study is based on a similar one, conducted by Ihorst (1993), Chapter 8.3. Let $\mathbf{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \mathbf{W}_{11})$ and $\mathbf{Y} \sim \mathcal{N}_2(\boldsymbol{\mu}, \mathbf{W}_{22})$ be two bivariate normally distributed random vectors with common mean $\boldsymbol{\mu}$. Further, let $\mathbf{X}$ and $\mathbf{Y}$ be correlated with $\mathrm{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{W}_{12}$, such that altogether we have a 4-variate normal distribution $(\mathbf{X}^\top, \mathbf{Y}^\top)^\top \sim \mathcal{N}_4((\boldsymbol{\mu}^\top, \boldsymbol{\mu}^\top)^\top, \mathbf{W})$ with

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{12}^\top & \mathbf{W}_{22} \end{pmatrix} .$$

Given a random sample $(\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_n, \mathbf{Y}_n)$ from $(\mathbf{X}, \mathbf{Y})$ we can calculate the arithmetic means $\mathbf{T}_1 = \overline{\mathbf{X}}$ and $\mathbf{T}_2 = \overline{\mathbf{Y}}$ from the respective subsamples as estimators for $\boldsymbol{\mu}$. For these estimators we have

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^\top & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \mathrm{Cov} \begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \end{pmatrix} = \mathrm{Cov} \begin{pmatrix} \overline{\mathbf{X}} \\ \overline{\mathbf{Y}} \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{12}^\top & \mathbf{W}_{22} \end{pmatrix} = \frac{1}{n}\mathbf{W} .$$

Our objective is to make use of the correlation between $\mathbf{T}_1$ and $\mathbf{T}_2$ in order to calculate linearly combined estimators employing the three covariance adjustment methods. Then these combined estimators are compared to each other as well as to the original estimators $\mathbf{T}_1$ and $\mathbf{T}_2$ and to the arithmetic mean $\mathbf{T}_{AM}$ of $\mathbf{T}_1$ and $\mathbf{T}_2$.
We investigated several choices of the covariance matrix $\mathbf{W}$ of $(\mathbf{X}^\top, \mathbf{Y}^\top)^\top$. In order to have some continuity with respect to previous work, we used the positive definite matrices from Ihorst (1993), Chapter 8.3 $(\mathbf{W}_1, \mathbf{W}_3, \mathbf{W}_4, \mathbf{W}_6, \mathbf{W}_7, \mathbf{W}_9, \mathbf{W}_{10})$. These were chosen from a set of randomly generated covariance matrices in such a way that the corresponding α_0 values (for WCAT optimal combination) cover the whole range of interesting constellations: There are values close to $1/2$, where you would expect the arithmetic mean to also deliver good results, values not so close to $1/2$ but still in the interval $[0, 1]$, and extreme values even outside $[0, 1]$ which indicate that the two individual estimators are very different in accuracy.
In order to provide an interesting range of optimal weight matrices $\mathbf{L}_0$ we have supplemented these choices by three further covariance matrices: $\mathbf{W}_2$ stands for the case where $\mathbf{L}_0$ is approximately diagonal, whereas $\mathbf{W}_5$ produces an exactly diagonal $\mathbf{L}_0$. With $\mathbf{W}_4$ the optimal weight is almost of the form $\alpha\mathbf{I}$, whereas $\mathbf{W}_8$ has the optimal weight exactly of this form. Another interesting point concerning $\mathbf{W}_8$ is the following: The optimal weight is $\mathbf{L}_0 = \mathbf{0}$, meaning that we should only use the single estimator $\mathbf{T}_1$ and neglect the information contained in $\mathbf{T}_2$. Of course this seems to be suboptimal, but Dickinson (1988) demonstrated that in the present situation no gain of accuracy (i.e. variance) can be achieved by a convex combination of $\mathbf{T}_1$ and $\mathbf{T}_2$. Namely, the situation corresponds to the univariate case, where the correlation coefficient $\rho = \sigma_{12}(\sigma_1\sigma_2)^{-1}$ between the two estimators equals the ratio of the respective standard deviations σ_1/σ_2, with $\sigma_1 < \sigma_2$.

In the following we list the covariance matrices under study and the corresponding optimal combination weights ordered by increasing $|\alpha_0 - 1/2|$:

$$\mathbf{W}_1 = \begin{pmatrix} 3 & -5 & -1 & -2 \\ -5 & 13 & 0 & -1 \\ -1 & 0 & 6 & 4 \\ -2 & -1 & 4 & 6 \end{pmatrix}$$
with $\mathrm{tr}(\mathbf{W}_1^\top \mathbf{W}_1) = 344$
$\alpha_0 = 0.5625$
$\mathbf{D}_0 = \mathrm{diag}(0.3636, 0.6667)$
$$\mathbf{L}_0 = \begin{pmatrix} 0.3783 & -0.1609 \\ -0.5174 & 0.6913 \end{pmatrix}$$
$\mathrm{tr}(\mathbf{L}_0^\top \mathbf{L}_0) = 0.9146$

$$\mathbf{W}_2 = \begin{pmatrix} 7 & 6 & -5 & -8 \\ 6 & 37 & -21 & -8 \\ -5 & -21 & 14 & 5 \\ -8 & -8 & 5 & 12 \end{pmatrix}$$
with $\mathrm{tr}(\mathbf{W}_2^\top \mathbf{W}_2) = 3068$
$\alpha_0 = 0.5938$
$\mathbf{D}_0 = \mathrm{diag}(0.3871, 0.6923)$
$$\mathbf{L}_0 = \begin{pmatrix} 0.5301 & -0.1108 \\ -0.1084 & 0.7590 \end{pmatrix}$$
$\mathrm{tr}(\mathbf{L}_0^\top \mathbf{L}_0) = 0.8812$

$$\mathbf{W}_3 = \begin{pmatrix} 7 & -6 & 3 & -4 \\ -6 & 18 & -9 & 12 \\ 3 & -9 & 25 & -7 \\ -4 & 12 & -7 & 10 \end{pmatrix}$$
with $\mathrm{tr}(\mathbf{W}_3^\top \mathbf{W}_3) = 1768$
$\alpha_0 = 0.3333$
$\mathbf{D}_0 = \mathrm{diag}(0.1538, 1.5000)$
$$\mathbf{L}_0 = \begin{pmatrix} 0.1538 & -0.5 \\ 0.1154 & 1.5 \end{pmatrix}$$
$\mathrm{tr}(\mathbf{L}_0^\top \mathbf{L}_0) = 2.537$

$$\mathbf{W}_4 = \begin{pmatrix} 18 & 10 & -6 & -1 \\ 10 & 19 & -5 & -2 \\ -6 & -5 & 6 & 5 \\ -1 & -2 & 5 & 6 \end{pmatrix}$$
with $\mathrm{tr}(\mathbf{W}_4^\top \mathbf{W}_4) = 1139$
$\alpha_0 = 0.6923$
$\mathbf{D}_0 = \mathrm{diag}(0.6667, 0.7241)$
$$\mathbf{L}_0 = \begin{pmatrix} 0.7711 & -0.1791 \\ -0.0100 & 0.7313 \end{pmatrix}$$
$\mathrm{tr}(\mathbf{L}_0^\top \mathbf{L}_0) = 1.1617$

$$\mathbf{W}_5 = \begin{pmatrix} 3 & 0 & 1 & 0 \\ 0 & 8 & 0 & 5 \\ 1 & 0 & 9 & 0 \\ 0 & 5 & 0 & 9 \end{pmatrix}$$
with $\mathrm{tr}(\mathbf{W}_5^\top \mathbf{W}_5) = 287$
$\alpha_0 = 0.2941$
$\mathbf{D}_0 = \mathrm{diag}(0.2000, 0.4286)$
$$\mathbf{L}_0 = \begin{pmatrix} 0.2000 & 0.0000 \\ 0.0000 & 0.4286 \end{pmatrix}$$
$\mathrm{tr}(\mathbf{L}_0^\top \mathbf{L}_0) = 0.2237$

$$\mathbf{W}_6 = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 2 & -2 & 0 \\ 1 & -2 & 4 & 1 \\ 1 & 0 & 1 & 3 \end{pmatrix}$$
with $\mathrm{tr}(\mathbf{W}_6^\top \mathbf{W}_6) = 44$
$\alpha_0 = 0.2500$
$\mathbf{D}_0 = \mathrm{diag}(0.0000, 0.4000)$
$$\mathbf{L}_0 = \begin{pmatrix} 0.1818 & -0.2727 \\ 0.5455 & 0.1818 \end{pmatrix}$$
$\mathrm{tr}(\mathbf{L}_0^\top \mathbf{L}_0) = 0.438$

$$\mathbf{W}_7 = \begin{pmatrix} 21 & 7 & 17 & 16 \\ 7 & 19 & 24 & 18 \\ 17 & 24 & 35 & 25 \\ 16 & 18 & 25 & 26 \end{pmatrix}$$
with $\mathrm{tr}(\mathbf{W}_7^\top \mathbf{W}_7) = 6941$
$\alpha_0 = 0.1613$
$\mathbf{D}_0 = \mathrm{diag}(0.1818, 0.1111)$
$\mathbf{L}_0 = \begin{pmatrix} -0.2687 & -1.2388 \\ -1.0821 & -0.8507 \end{pmatrix}$
$\mathrm{tr}(\mathbf{L}_0^\top \mathbf{L}_0) = 3.5015$

$$\mathbf{W}_8 = \begin{pmatrix} 4 & 0 & 4 & 0 \\ 0 & 4 & 0 & 4 \\ 4 & 0 & 7 & 0 \\ 0 & 4 & 0 & 7 \end{pmatrix}$$
with $\mathrm{tr}(\mathbf{W}_8^\top \mathbf{W}_8) = 194$
$\alpha_0 = 0$
$\mathbf{D}_0 = \mathrm{diag}(0.0000, 0.0000)$
$\mathbf{L}_0 = \begin{pmatrix} 0.0000 & 0.0000 \\ 0.0000 & 0.0000 \end{pmatrix}$
$\mathrm{tr}(\mathbf{L}_0^\top \mathbf{L}_0) = 0$

$$\mathbf{W}_9 = \begin{pmatrix} 3 & -1 & 6 & -2 \\ -1 & 1 & -2 & 1 \\ 6 & -2 & 15 & 2 \\ -2 & 1 & 2 & 27 \end{pmatrix}$$
with $\mathrm{tr}(\mathbf{W}_9^\top \mathbf{W}_9) = 1064$
$\alpha_0 = -0.0938$
$\mathbf{D}_0 = \mathrm{diag}(-0.5000, 0.0000)$
$\mathbf{L}_0 = \begin{pmatrix} -0.6336 & 0.1603 \\ 0.1985 & -0.0382 \end{pmatrix}$
$\mathrm{tr}(\mathbf{L}_0^\top \mathbf{L}_0) = 0.468$

$$\mathbf{W}_{10} = \begin{pmatrix} 18 & 15 & 17 & 11 \\ 15 & 18 & 16 & 13 \\ 17 & 16 & 18 & 11 \\ 11 & 13 & 11 & 10 \end{pmatrix}$$
with $\mathrm{tr}(\mathbf{W}_{10}^\top \mathbf{W}_{10}) = 3434$
$\alpha_0 = 1.5000$
$\mathbf{D}_0 = \mathrm{diag}(0.5000, 2.5000)$
$\mathbf{L}_0 = \begin{pmatrix} 2.0000 & 3.0000 \\ 1.0000 & 3.0000 \end{pmatrix}$
$\mathrm{tr}(\mathbf{L}_0^\top \mathbf{L}_0) = 23$.

Since the joint covariance matrix $\mathbf{\Sigma}$ of $\mathbf{T}_1$ and $\mathbf{T}_2$ is assumed to be unknown, it has to be estimated in order to apply the covariance adjustment techniques. For each of the choices for $\mathbf{W}$ we drew three random samples of different sizes n from the corresponding 4-variate normal distribution. Without loss of generality we chose $\boldsymbol{\mu} = \mathbf{0}$, i.e. we drew samples $(\mathbf{X}_1^\top, \mathbf{Y}_1^\top)^\top, \ldots, (\mathbf{X}_n^\top, \mathbf{Y}_n^\top)^\top$ from $(\mathbf{X}^\top, \mathbf{Y}^\top)^\top \sim \mathcal{N}_4(\mathbf{0}, \mathbf{W})$. The respective sample sizes were $n = 10$, $n = 25$ and $n = 50$.
Using these random data we calculated the arithmetic means

$$\mathbf{T}_1 = \overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \qquad \text{and} \qquad \mathbf{T}_2 = \overline{\mathbf{Y}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y}_i$$

from the $\mathbf{X}$- and $\mathbf{Y}$-samples as estimates for $\boldsymbol{\mu} = \mathbf{0}$.
To estimate $\mathbf{\Sigma}$ we employed the sample covariance matrix $\widehat{\mathbf{W}}$:

$$\begin{aligned} \widehat{\mathbf{\Sigma}} &= \begin{pmatrix} \widehat{\mathbf{\Sigma}}_{11} & \widehat{\mathbf{\Sigma}}_{12} \\ \widehat{\mathbf{\Sigma}}_{12}^\top & \widehat{\mathbf{\Sigma}}_{22} \end{pmatrix} = \frac{1}{n}\begin{pmatrix} \widehat{\mathbf{W}}_{11} & \widehat{\mathbf{W}}_{12} \\ \widehat{\mathbf{W}}_{12}^\top & \widehat{\mathbf{W}}_{22} \end{pmatrix} = \frac{1}{n}\widehat{\mathbf{W}} \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\begin{pmatrix} \mathbf{X}_i \\ \mathbf{Y}_i \end{pmatrix} - \begin{pmatrix} \overline{\mathbf{X}} \\ \overline{\mathbf{Y}} \end{pmatrix} \right) \left(\begin{pmatrix} \mathbf{X}_i \\ \mathbf{Y}_i \end{pmatrix} - \begin{pmatrix} \overline{\mathbf{X}} \\ \overline{\mathbf{Y}} \end{pmatrix} \right)^\top . \end{aligned}$$

Using this estimator we can calculate estimates for the combination weights of the respective covariance adjustment techniques.
$\widehat{\text{SCAT}}$: Estimate $\mathbf{L}_0$ by

$$\widehat{\mathbf{L}}_0 = (\widehat{\mathbf{\Sigma}}_{11} - \widehat{\mathbf{\Sigma}}_{12})(\widehat{\mathbf{\Sigma}}_{11} + \widehat{\mathbf{\Sigma}}_{22} - \widehat{\mathbf{\Sigma}}_{12} - \widehat{\mathbf{\Sigma}}_{12}^{\top})^{-1} .$$

$\widehat{\text{MCAT}}$: Estimate $\mathbf{D}_0$ by

$$\widehat{\mathbf{D}}_0 = \text{diag}(\widehat{\mathbf{\Sigma}}_{11} - \widehat{\mathbf{\Sigma}}_{12})(\text{diag}(\widehat{\mathbf{\Sigma}}_{11} + \widehat{\mathbf{\Sigma}}_{22} - 2\widehat{\mathbf{\Sigma}}_{12}))^{-1} .$$

$\widehat{\text{WCAT}}$: Estimate α_0 by

$$\widehat{\alpha}_0 = \text{tr}(\widehat{\mathbf{\Sigma}}_{11} - \widehat{\mathbf{\Sigma}}_{12})(\text{tr}(\widehat{\mathbf{\Sigma}}_{11} + \widehat{\mathbf{\Sigma}}_{22} - 2\widehat{\mathbf{\Sigma}}_{12}))^{-1} .$$

As already mentioned in Section 2 in practice we can observe that estimating the optimal combination weights neglecting covariances between the estimators often leads to better combined estimates than calculation employing these covariances (cf. Makridakis and Winkler (1983)). Therefore, we will also investigate variants of the three covariance adjustment techniques neglecting covariances between estimators, i.e. assuming them to be $\mathbf{0}$, in the calculation of the combination weights. These variants will be referred to as $\widetilde{\text{SCAT}}$, $\widetilde{\text{MCAT}}$ and $\widetilde{\text{WCAT}}$, respectively.
$\widetilde{\text{SCAT}}$: Estimate $\mathbf{L}_0$ by

$$\widetilde{\mathbf{L}}_0 = \widehat{\mathbf{\Sigma}}_{11}(\widehat{\mathbf{\Sigma}}_{11} + \widehat{\mathbf{\Sigma}}_{22})^{-1} .$$

$\widetilde{\text{MCAT}}$: Estimate $\mathbf{D}_0$ by

$$\widetilde{\mathbf{D}}_0 = \text{diag}(\widehat{\mathbf{\Sigma}}_{11})(\text{diag}(\widehat{\mathbf{\Sigma}}_{11} + \widehat{\mathbf{\Sigma}}_{22}))^{-1} .$$

$\widetilde{\text{WCAT}}$: Estimate α_0 by

$$\widetilde{\alpha}_0 = \text{tr}(\widehat{\mathbf{\Sigma}}_{11})(\text{tr}(\widehat{\mathbf{\Sigma}}_{11} + \widehat{\mathbf{\Sigma}}_{22}))^{-1} .$$

By plugging in these combination weight estimators we obtain the parameter estimators $\mathbf{T}_{\widehat{\alpha}_0}$, $\mathbf{T}_{\widetilde{\alpha}_0}$, $\mathbf{T}_{\widehat{\mathbf{D}}_0}$, $\mathbf{T}_{\widetilde{\mathbf{D}}_0}$, $\mathbf{T}_{\widehat{\mathbf{L}}_0}$ and $\mathbf{T}_{\widetilde{\mathbf{L}}_0}$, e.g. $\mathbf{T}_{\widehat{\mathbf{L}}_0} = (\mathbf{I} - \widehat{\mathbf{L}}_0)\mathbf{T}_1 + \widehat{\mathbf{L}}_0\mathbf{T}_2$.
The whole process of drawing random samples and calculating estimates is repeated 1 000 times. To judge the performance of an estimator $\mathbf{T}$ for $\boldsymbol{\mu} = \mathbf{0}$ we calculate the average of the sum of squared errors $\text{av}\,(\text{SSE}(\mathbf{T}, \boldsymbol{\mu}))$ over the 1 000 simulation runs, where

$$\text{SSE}(\mathbf{T}, \boldsymbol{\mu}) = (\mathbf{T} - \boldsymbol{\mu})^{\top}(\mathbf{T} - \boldsymbol{\mu}) = \mathbf{T}^{\top}\mathbf{T} .$$

Thus, an estimator will be called *better* than another if it has a smaller average SSE-value.

As mentioned earlier the optimal SCAT-combination weight $\mathbf{L}_0$ is determined in order to minimize the covariance matrix of the combined estimator in the sense of the Löwner ordering. On the other hand the optimal WCAT-combination weight α_0 is determined in order to minimize the total variance of the combined estimator. Furthermore, if an estimator is better than another with respect to the covariance matrix criterion it is also better with respect to the total variance criterion and the total variance criterion coincides with the scalar mean square error for unbiased estimators. Consequently, the average SSE-value is the natural choice for the performance measure in this simulation study, since it is the empirical counterpart of the SMSE:

$$\mathrm{SMSE}(\mathbf{T}, \boldsymbol{\mu}) = \mathrm{E}\left((\mathbf{T} - \boldsymbol{\mu})^\top (\mathbf{T} - \boldsymbol{\mu})\right) .$$

Being a real number is a further advantage of the average SSE-value, since comparisons with respect to this measure are easily done. Moreover, we can determine by what percentage one estimator outperforms another.

In the simulation study the covariance matrix $\mathbf{W}$ of $(\mathbf{X}^\top, \mathbf{Y}^\top)^\top$ and, hence, the optimal combination weights are given. Therefore, we can also calculate the CAT estimators using the respective *optimal* combination weights $\mathbf{L}_0$, $\mathbf{D}_0$ and $\alpha_0\mathbf{I}$. The corresponding average SSE-values may serve as an indicator how good the CAT estimators using the *estimated* combination weights may get.

Tables 1 and 2 report the average SSE-values of the respective estimates for the parameter vector $\boldsymbol{\mu} = \mathbf{0}$ relative to the average SSE-value that is obtained for the theoretically best estimate $\mathbf{T}_{\mathbf{L}_0}$, i.e. for any estimator $\mathbf{T}$ we report

$$\frac{\mathrm{av}\left(\mathrm{SSE}(\mathbf{T}, \boldsymbol{\mu})\right)}{\mathrm{av}\left(\mathrm{SSE}(\mathbf{T}_{\mathbf{L}_0}, \boldsymbol{\mu})\right)} = \frac{\mathrm{av}\left(\mathbf{T}^\top \mathbf{T}\right)}{\mathrm{av}\left(\mathbf{T}_{\mathbf{L}_0}^\top \mathbf{T}_{\mathbf{L}_0}\right)} .$$

These values have been truncated after the second decimal. It should be remarked that we will denote the relative SSE-values for those estimators which are exactly as good as $\mathbf{T}_{\mathbf{L}_0}$ as '1', whereas we will denote '1.00' for those estimators which perform equally good as $\mathbf{T}_{\mathbf{L}_0}$ within the tolerance of this table, i.e. which produce SSE-values between 1.000 and 1.009.

The SSE-values of the combinations which depend on fixed weights ($\mathbf{T}_1$, $\mathbf{T}_2$, $\mathbf{T}_{\alpha_0}$, $\mathbf{T}_{\mathbf{D}_0}$, $\mathbf{T}_{\mathbf{L}_0}$, $\mathbf{T}_{AM}$) could have been calculated theoretically, i.e. without simulation. Whenever combination weights have to be estimated, however, direct calculation of the errors is not possible and simulation techniques have to be used. For reasons of the homogeneity of the reported values the tables present the simulation values throughout.

Furthermore, we can calculate the goodness of the estimated covariance matrix $\widehat{\mathbf{\Sigma}}$ with respect to the true covariance matrix $\mathbf{\Sigma}$, and we can calculate the goodness of the estimated combination weights with respect to the optimal weight matrix $\mathbf{L}_0$ as well. Again this is done by the averages of the respective sums of squared errors, i.e. by the averages of the squared Frobenius norms (cf. Horn and Johnson (1985)):

$$\text{av}\left(\text{tr}\left((\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma})^\top(\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma})\right)\right) \;;$$

$$\text{av}\left(\text{tr}\left((\widehat{\mathbf{L}}_0-\mathbf{L}_0)^\top(\widehat{\mathbf{L}}_0-\mathbf{L}_0)\right)\right) \;, \quad \text{av}\left(\text{tr}\left((\widetilde{\mathbf{L}}_0-\mathbf{L}_0)^\top(\widetilde{\mathbf{L}}_0-\mathbf{L}_0)\right)\right) \;,$$

$$\text{av}\left(\text{tr}\left((\widehat{\mathbf{D}}_0-\mathbf{L}_0)^\top(\widehat{\mathbf{D}}_0-\mathbf{L}_0)\right)\right) \;, \quad \text{av}\left(\text{tr}\left((\widetilde{\mathbf{D}}_0-\mathbf{L}_0)^\top(\widetilde{\mathbf{D}}_0-\mathbf{L}_0)\right)\right)$$

and

$$\text{av}\left(\text{tr}\left((\widehat{\alpha}_0\mathbf{I}-\mathbf{L}_0)^\top(\widehat{\alpha}_0\mathbf{I}-\mathbf{L}_0)\right)\right) \;, \quad \text{av}\left(\text{tr}\left((\widetilde{\alpha}_0\mathbf{I}-\mathbf{L}_0)^\top(\widetilde{\alpha}_0\mathbf{I}-\mathbf{L}_0)\right)\right) \;.$$

Table 3 gives the average SSE-values for the estimation of the covariance matrix $\mathbf{\Sigma}$ and for the estimation of the combination weights, respectively. These SSE-values have been divided by the respective sum of squares of the corresponding true or optimal matrices, i.e. we list

$$\frac{\text{av}\left(\text{tr}\left((\widehat{\mathbf{W}}-\mathbf{W})^\top(\widehat{\mathbf{W}}-\mathbf{W})\right)\right)}{\text{tr}\left(\mathbf{W}^\top\mathbf{W}\right)} \qquad \left(=\frac{\text{av}\left(\text{tr}\left((\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma})^\top(\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma})\right)\right)}{\text{tr}\left(\mathbf{\Sigma}^\top\mathbf{\Sigma}\right)}\right)$$

in the second column and

$$\frac{\text{av}\left(\text{tr}\left((\mathbf{M}-\mathbf{L}_0)^\top(\mathbf{M}-\mathbf{L}_0)\right)\right)}{\text{tr}\left(\mathbf{L}_0^\top\mathbf{L}_0\right)}$$

in the following six columns, where $\mathbf{M}$ denotes the respective estimate for the combination weight. All these values have been rounded to the fourth decimal. As an exception, with covariance matrix $\mathbf{W}_8$ we could not divide by the sum of squares of $\mathbf{L}_0$, since $\mathbf{L}_0 = \mathbf{0}$. Here, in columns 3 through 8 we report the average SSE-values $\text{av}\left(\text{tr}\left((\mathbf{M}-\mathbf{L}_0)^\top(\mathbf{M}-\mathbf{L}_0)\right)\right)$ without normation by $\text{tr}\left(\mathbf{L}_0^\top\mathbf{L}_0\right)$ instead. This is indicated by a '$***$' in the corresponding section of Table 3.

Table 1: Estimation of parameter vector $\boldsymbol{\mu} = \mathbf{0}$: Sum of squared errors (average from 1000 simulation runs) relative to sum of squared errors of $\mathbf{T}_{\mathbf{L}_0}$ (average from 1000 simulation runs)

	single estimators		arithmetic mean	WCAT estimators			
$\mathbf{W}_i$	$\mathbf{T}_1$	$\mathbf{T}_2$	$\mathbf{T}_{AM}$	$\mathbf{T}_{\alpha_0}$	$\mathbf{T}_{\widehat{\alpha_0}}$	$\mathbf{T}_{\widetilde{\alpha_0}}$	n
$\mathbf{W}_1$	8.73	6.71	3.29	3.23	3.59	3.65	10
	10.00	6.87	3.58	3.46	3.57	3.60	25
	9.00	6.51	3.36	3.27	3.28	3.30	50
$\mathbf{W}_2$	5.73	3.31	1.42	1.30	1.33	1.35	10
	5.67	3.30	1.43	1.32	1.33	1.34	25
($\mathbf{L}_0 \approx$ diag)	5.81	3.54	1.34	1.25	1.26	1.28	50
$\mathbf{W}_3$	1.71	2.57	1.60	1.52	1.58	1.55	10
	1.74	2.33	1.56	1.51	1.53	1.53	25
	1.85	2.44	1.62	1.58	1.60	1.59	50
$\mathbf{W}_4$	6.58	2.14	1.55	1.11	1.19	1.25	10
	7.47	2.29	1.64	1.12	1.16	1.21	25
($\mathbf{L}_0 \approx \alpha\mathbf{I}_2$)	7.92	2.29	1.71	1.13	1.15	1.18	50
$\mathbf{W}_5$	1.18	1.92	1.09	1.02	1.06	1.04	10
	1.13	2.06	1.12	1.01	1.03	1.04	25
($\mathbf{L}_0 =$ diag)	1.20	1.90	1.08	1.01	1.03	1.03	50
$\mathbf{W}_6$	2.46	5.97	2.48	2.04	2.12	2.10	10
	2.36	5.37	2.29	1.93	1.98	1.96	25
	2.25	5.16	2.22	1.87	1.88	1.88	50
$\mathbf{W}_7$	3.37	5.34	3.70	3.33	3.54	3.54	10
	3.19	4.89	3.41	3.13	3.22	3.28	25
	3.12	4.87	3.38	3.07	3.11	3.23	50
$\mathbf{W}_8$	1	1.80	1.20	1	1.06	1.11	10
	1	1.70	1.16	1	1.02	1.09	25
($\mathbf{L}_0 = \alpha\mathbf{I}_2$)	1	1.73	1.19	1	1.00	1.10	50
$\mathbf{W}_9$	2.44	25.92	9.30	2.24	2.40	2.96	10
	2.26	24.56	8.67	2.12	2.14	2.72	25
	2.28	25.48	8.90	2.14	2.17	2.72	50
$\mathbf{W}_{10}$	4.45	3.52	3.85	3.44	3.66	3.80	10
	4.57	3.49	3.90	3.33	3.40	3.84	25
	4.55	3.52	3.91	3.38	3.41	3.85	50

Table 2: Estimation of parameter vector $\boldsymbol{\mu} = \mathbf{0}$: Sum of squared errors (average from 1000 simulation runs) relative to sum of squared errors of $\mathbf{T}_{\mathbf{L}_0}$ (average from 1000 simulation runs) (continued)

	MCAT estimators			SCAT estimators			
$\mathbf{W}_i$	$\mathbf{T}_{\mathbf{D}_0}$	$\mathbf{T}_{\widehat{\mathbf{D}}_0}$	$\mathbf{T}_{\widetilde{\mathbf{D}}_0}$	$\mathbf{T}_{\mathbf{L}_0}$	$\mathbf{T}_{\widehat{\mathbf{L}}_0}$	$\mathbf{T}_{\widetilde{\mathbf{L}}_0}$	n
$\mathbf{W}_1$	2.86	3.35	3.33	1	1.40	1.46	10
	3.03	3.17	3.19	1	1.09	1.25	25
	2.92	2.91	2.94	1	1.02	1.16	50
$\mathbf{W}_2$	1.01	1.17	1.19	1	1.28	1.22	10
	1.03	1.06	1.10	1	1.07	1.09	25
($\mathbf{L}_0 \approx$ diag)	1.02	1.05	1.12	1	1.04	1.06	50
$\mathbf{W}_3$	1.09	1.24	1.31	1	1.29	1.25	10
	1.10	1.13	1.30	1	1.07	1.23	25
	1.11	1.14	1.36	1	1.04	1.28	50
$\mathbf{W}_4$	1.10	1.24	1.28	1	1.32	1.39	10
	1.11	1.16	1.21	1	1.08	1.20	25
($\mathbf{L}_0 \approx \alpha\mathbf{I}_2$)	1.12	1.15	1.19	1	1.05	1.18	50
$\mathbf{W}_5$	1	1.11	1.02	1	1.32	1.06	10
	1	1.04	1.02	1	1.09	1.03	25
($\mathbf{L}_0 =$ diag)	1	1.03	1.01	1	1.03	1.00	50
$\mathbf{W}_6$	1.76	2.10	2.08	1	1.38	2.22	10
	1.66	1.76	1.83	1	1.11	1.92	25
	1.65	1.69	1.76	1	1.04	1.84	50
$\mathbf{W}_7$	3.34	3.84	3.57	1	1.45	3.42	10
	3.13	3.30	3.29	1	1.10	3.09	25
	3.07	3.14	3.23	1	1.04	3.01	50
$\mathbf{W}_8$	1	1.16	1.12	1	1.36	1.13	10
	1	1.04	1.09	1	1.09	1.09	25
($\mathbf{L}_0 = \alpha\mathbf{I}_2$)	1	1.01	1.10	1	1.03	1.10	50
$\mathbf{W}_9$	1.44	1.63	3.24	1	1.33	3.48	10
	1.43	1.49	2.95	1	1.09	3.08	25
	1.40	1.44	3.01	1	1.03	3.11	50
$\mathbf{W}_{10}$	2.88	3.34	3.72	1	1.38	3.34	10
	2.90	3.08	3.77	1	1.08	3.43	25
	2.90	2.96	3.77	1	1.05	3.43	50

Table 3: Estimation of covariance matrix and combination weights: Sum of squared errors (average from 1000 simulation runs) relative to the FROBENIUS norms of $\mathbf{W}$ and $\mathbf{L}_0$, respectively

	error vs. $\mathbf{W}$	errors vs. $\mathbf{L}_0$						
$\mathbf{W}_i$	$\hat{\mathbf{W}}$	$\widehat{\alpha}_0\mathbf{I}_2$	$\widetilde{\alpha}_0\mathbf{I}_2$	$\widehat{\mathbf{D}}_0$	$\widetilde{\mathbf{D}}_0$	$\widehat{\mathbf{L}}_0$	$\widetilde{\mathbf{L}}_0$	n
$\mathbf{W}_1$	0.3766	0.4160	0.4219	0.3751	0.3734	0.0429	0.0585	10
	0.1408	0.3894	0.3929	0.3379	0.3401	0.0130	0.0296	25
	0.0647	0.3825	0.3851	0.3290	0.3312	0.0059	0.0225	50
$\mathbf{W}_2$	0.3311	0.0710	0.0685	0.0852	0.1038	0.3149	0.1776	10
	0.1135	0.0654	0.0611	0.0630	0.0800	0.0973	0.1164	25
($\mathbf{L}_0 \approx$ diag)	0.0520	0.0641	0.0594	0.0591	0.0751	0.0486	0.0994	50
$\mathbf{W}_3$	0.3173	0.6721	0.6026	0.2425	0.4050	0.2574	0.3269	10
	0.1348	0.6541	0.5945	0.1520	0.3984	0.0793	0.3161	25
	0.0618	0.6572	0.5961	0.1249	0.3967	0.0345	0.3146	50
$\mathbf{W}_4$	0.3481	0.0475	0.0442	0.0612	0.0528	0.0855	0.0862	10
	0.1278	0.0395	0.0344	0.0446	0.0371	0.0218	0.0379	25
($\mathbf{L}_0 \approx \alpha\mathbf{I}_2$)	0.0620	0.0364	0.0312	0.0404	0.0326	0.0107	0.0288	50
$\mathbf{W}_5$	0.4493	0.4185	0.2560	0.7680	0.1947	1.6921	0.3396	10
	0.1674	0.2216	0.1898	0.2474	0.0808	0.4837	0.1313	25
($\mathbf{L}_0 =$ diag)	0.0804	0.1683	0.1736	0.1165	0.0493	0.2161	0.0706	50
$\mathbf{W}_6$	0.3869	0.9224	0.9416	1.2176	1.0518	0.3229	1.2786	10
	0.1393	0.8919	0.9277	1.0963	0.9989	0.1051	1.1881	25
	0.0681	0.8792	0.9179	1.0567	0.9719	0.0473	1.1667	50
$\mathbf{W}_7$	0.2695	1.1430	1.3550	1.2338	1.3663	0.1534	1.2968	10
	0.0999	1.1215	1.3480	1.1293	1.3577	0.0394	1.2763	25
	0.0481	1.1183	1.3446	1.1120	1.3557	0.0173	1.2747	50
$\mathbf{W}_8$	0.3848	0.1557	0.2754	0.3483	0.2919	0.8472	0.3178	10
$* * *$	0.1473	0.0608	0.2707	0.1200	0.2766	0.2522	0.2832	25
($\mathbf{L}_0 = \alpha\mathbf{I}_2$)	0.0682	0.0278	0.2639	0.0556	0.2663	0.1139	0.2690	50
$\mathbf{W}_9$	0.3304	0.7768	1.2979	0.2638	1.5461	0.1469	1.7344	10
	0.1232	0.7630	1.2882	0.2077	1.5246	0.0420	1.6424	25
	0.0575	0.7699	1.2872	0.1924	1.5234	0.0193	1.6306	50
$\mathbf{W}_{10}$	0.1947	0.5655	0.7831	0.6122	0.7750	0.0809	0.7289	10
	0.0944	0.5507	0.7832	0.5695	0.7745	0.0226	0.7264	25
	0.0458	0.5448	0.7828	0.5522	0.7742	0.0101	0.7272	50

4 Results

What is the best choice? The best technique for every combination of covariance matrix $\mathbf{W}$ and sample size n is SCAT employing the optimal weight. This justifies our decision to present all the SSE-values relative to the SSE-value of $\mathbf{T}_{\mathbf{L}_0}$.

Optimal vs. estimated combination weights: As could be expected the covariance adjustment techniques using the optimal combination weights, $\mathbf{T}_{\alpha_0}, \mathbf{T}_{\mathbf{D}_0}$ and $\mathbf{T}_{\mathbf{L}_0}$, almost always outperform the respective variants where the combination weights are estimated. In empirical situations, however, the true covariance structure is unknown, so that the CAT variants using the optimal weights are not feasible. Consequently, the interesting CAT variants are these, where the combination weights have to be estimated, and we will regard only them from now on.

A rough overall judgement: In general the $\widehat{\text{SCAT}}$ variant, i.e. $\mathbf{T}_{\widehat{\mathbf{L}}_0}$, is the best estimation procedure by far. It almost always outperforms both of the original estimators $\mathbf{T}_1$ and $\mathbf{T}_2$ (exceptions ($\mathbf{W}_5$, $n = 10$) and ($\mathbf{W}_8$, all n)). This could have been hoped for, since it is a combined estimator using much more information than the original estimators. Whenever the optimal weight $\mathbf{L}_0$ is equal to or close to a diagonal matrix, other combined estimators may exhibit a slight advantage (see later), but otherwise $\mathbf{T}_{\widehat{\mathbf{L}}_0}$ outperforms all of the other combined estimators, CAT variants as well as the arithmetic mean. Very often the improvement gained by using $\widehat{\text{SCAT}}$ is enormous, with relative SSE-values which are better than those of the alternatives up to 200%. Sometimes the gain is even larger. And *if* any of the alternatives outperforms $\mathbf{T}_{\widehat{\mathbf{L}}_0}$ it is only by a small margin. Hence, use of this SCAT variant is strictly recommended.

For MCAT and WCAT the $\hat{}$-variant and the $\tilde{}$-variant perform almost equally well (exceptions $\mathbf{W}_9$ and $\mathbf{W}_{10}$, where the $\tilde{}$-variants are clearly worse). MCAT performs slightly better than WCAT in general with the exception of those covariance matrices $\mathbf{W}$ generating an $\mathbf{L}_0$ close to a multiple of the identity matrix. WCAT in turn performs slightly better than the arithmetic mean. (Exception: For $\mathbf{W}_9$, where the individual estimators $\mathbf{T}_1$ and $\mathbf{T}_2$ are very different in quality, the arithmetic mean is much worse than the other combined estimators.) All the combined estimators could outperform both individual estimators, with some exceptions for extreme α_0-values, where especially the $\tilde{}$-variants and the arithmetic mean ranked between the individual estimators. Any combination technique outperformed at least one of the original estimators.

Special optimal weight matrices $\mathbf{L}_0$: We will now consider the covariance matrices $\mathbf{W}$ which produce a special structure of the optimal weight matrix $\mathbf{L}_0$, namely:

$\mathbf{W}_2$ with $\mathbf{L}_0 \approx$ diag, $\mathbf{W}_5$ with $\mathbf{L}_0 =$ diag, $\mathbf{W}_4$ with $\mathbf{L}_0 \approx \alpha\mathbf{I}$, $\mathbf{W}_8$ with $\mathbf{L}_0 = \alpha\mathbf{I}$.

For $\mathbf{W}_2$ $\widehat{\text{MCAT}}$ gives very good results throughout. $\widetilde{\text{SCAT}}$ performs comparably good for all considered n, and so does $\widehat{\text{SCAT}}$ for $n = 25$ and $n = 50$. For $\mathbf{W}_5$ $\widetilde{\text{MCAT}}$ is a very good choice, $\widetilde{\text{SCAT}}$ and both WCAT variants are good alternatives; for $n = 25$ $\widehat{\text{MCAT}}$ and for $n = 50$ $\widehat{\text{MCAT}}$ and $\widehat{\text{SCAT}}$ join this *good group*. For ($\mathbf{W}_4$, $n = 10$) $\widehat{\text{WCAT}}$ is the best choice, but for $n = 25$ and $n = 50$ $\widehat{\text{SCAT}}$ is. For $\mathbf{W}_8$ $\widehat{\text{WCAT}}$ always holds the top position, for $n = 25$ $\widehat{\text{MCAT}}$ and for $n = 50$ $\widehat{\text{MCAT}}$ and $\widehat{\text{SCAT}}$ are comparable.

In general we can observe that the larger the sample size n the more we can trust on the $\widehat{\text{SCAT}}$ combination technique. In the case of smaller sample sizes we find evidence for the conjecture, that MCAT variants perform especially well when $\mathbf{L}_0$ is equal or close to diagonal, and that WCAT variants are very good when $\mathbf{L}_0$ is equal or close to a multiple of the identity matrix. Here we can state an advantage for the techniques employing the covariances between the parameter estimators, i.e. for the $\hat{}$-techniques.

Other optimal weight matrices: The remaining choices of $\mathbf{W}$ ($\mathbf{W}_1$, $\mathbf{W}_3$, $\mathbf{W}_6$, $\mathbf{W}_7$, $\mathbf{W}_9$ and $\mathbf{W}_{10}$) imply no special structure of $\mathbf{L}_0$. Presumably this will be the more relevant case for practical purposes. Here we can make a clear statement in favor of $\widehat{\text{SCAT}}$: It is the best combination technique, far ahead of the other techniques in most of the considered settings. Only for the less extreme values of α_0 and $n = 10$ $\widetilde{\text{SCAT}}$ can compete.

$\hat{}$-variants vs. $\tilde{}$-variants: We can state that $\widehat{\text{SCAT}}$ performs better than $\widetilde{\text{SCAT}}$ in general. But for some choices of $\mathbf{W}$ (especially those producing an almost diagonal $\mathbf{L}_0$) we can often observe the pattern that the $\tilde{}$-variant is better by a small margin for $n = 10$ and the $\hat{}$-variant is better by a small margin from $n = 25$ on. For most of the more extreme α_0-values ($\mathbf{W}_6$, $\mathbf{W}_7$, $\mathbf{W}_9$ and $\mathbf{W}_{10}$) the $\tilde{}$-variant exhibits very bad results. For WCAT and MCAT the $\hat{}$- and $\tilde{}$-variants are about equal throughout with slight advantages for the $\hat{}$-variant, especially for the larger n-values and for the very extreme α_0-values ($\mathbf{W}_9$ and $\mathbf{W}_{10}$).

We sometimes find that $\widetilde{\text{MCAT}}$ is slightly better than $\widehat{\text{MCAT}}$ for $n = 10$. Since MCAT is equivalent to the univariate treatment of the components of the parameter estimators, this fact to a certain extent supports the findings of Makridakis and Winkler (1983), that covariances between forecasts should be neglected. Makridakis and Winkler (1983) frequently used only small sample sizes.

How much is lost by the necessity to estimate? If we replace the SCAT optimal combination weight $\mathbf{L}_0$ by its estimate $\widehat{\mathbf{L}}_0$ employing covariances between the parameter estimators, we observe the following pattern for all choices of $\mathbf{W}$: For sample size $n = 10$ the estimator $\mathbf{T}_{\widehat{\mathbf{L}}_0}$ is about 30–40% worse than $\mathbf{T}_{\mathbf{L}_0}$, for

$n = 25$ we only lose about 10% and for $n = 50$ only about 4%, showing that $\mathbf{T}_{\widehat{\mathbf{L}}_0}$ is a good choice and that it tends to reproduce the results of the optimal method for large sample size n. The almost fixed losses of the $\widehat{\text{SCAT}}$ variant with respect to the optimal $\mathbf{T}_{\mathbf{L}_0}$ are a further advantage of $\widehat{\text{SCAT}}$, since the risk of using this technique can be estimated in advance.

When replacing $\mathbf{L}_0$ by its estimate $\widetilde{\mathbf{L}}_0$ neglecting covariances between the parameter estimators the results are not so uniform. Since covariances are neglected, the weight estimate converges to a wrong value, and hence the corresponding parameter estimate $\mathbf{T}_{\widetilde{\mathbf{L}}_0}$ may exhibit a loss of up to 250% especially when $|\alpha_0 - 1/2|$ is large. When $|\alpha_0 - 1/2|$ is small, however, the loss is also small and for $n = 10$ the loss of the $\tilde{}$-variant may even be smaller than that of the $\hat{}$-variant.

Taking $\mathbf{T}_{\widehat{\mathbf{D}}_0}$ instead of $\mathbf{T}_{\mathbf{D}_0}$ results in losses of about 10–20%, 5% and 3% for $n = 10$, $n = 25$ and $n = 50$, respectively. The losses with respect to $\mathbf{T}_{\mathbf{L}_0}$ differ widely for the various choices of the covariance matrix $\mathbf{W}$ as could be seen above (cf. the paragraphs on special / other optimal weight matrices). The losses for $\mathbf{T}_{\widetilde{\mathbf{D}}_0}$ are of about the same size except for the very extreme α_0-values ($\mathbf{W}_9$ and $\mathbf{W}_{10}$), where they are greater.

Using $\mathbf{T}_{\widehat{\alpha}_0}$ instead of $\mathbf{T}_{\alpha_0}$ results in losses of about 4–10%, 1–3% and 1% for $n = 10$, $n = 25$ and $n = 50$, respectively. Again the losses with respect to $\mathbf{T}_{\mathbf{L}_0}$ differ widely for the various choices of $\mathbf{W}$. The losses for $\mathbf{T}_{\widetilde{\alpha}_0}$ are of about the same size, again except for $\mathbf{W}_9$ and $\mathbf{W}_{10}$, where they are greater.

What is the effect of the sample size n? The general observation that the quality of parameter estimation improves with increasing sample size and that it approaches the optimal $\mathbf{T}_{\mathbf{L}_0}$ can only be stated for $\widehat{\text{SCAT}}$. For all other techniques the estimated combination weight appears to converge (cf. Table 3), but not to the optimal $\mathbf{L}_0$. Hence, parameter estimation remains suboptimal. Only for those choices of $\mathbf{W}$ with $\mathbf{L}_0 = \mathbf{D}_0$ ($\mathbf{L}_0 = \alpha_0\mathbf{I}_2$) MCAT (MCAT and WCAT) will approach the optimum with increasing n. What can be achieved in general can be seen from the columns for $\mathbf{T}_{\alpha_0}, \mathbf{T}_{\mathbf{D}_0}$ and $\mathbf{T}_{\mathbf{L}_0}$.

For $n = 50$ $\widehat{\text{SCAT}}$ produced very good results close to the optimal $\mathbf{T}_{\mathbf{L}_0}$, thus ranking first among the combination techniques. For $n = 25$ this dominance is slightly shattered ($\mathbf{W}_2$, $\mathbf{W}_5$, $\mathbf{W}_8$) and for $n = 10$ it is valid only for the covariance matrices effecting an optimal weight $\mathbf{L}_0$ not close to a diagonal matrix.

It should be pointed out that $n = 10$ is a very important choice for practical purposes: When applying the combination techniques to forecasts, the combination weights are frequently estimated by small samples like $n = 10$. This is done because in this context the relative quality of the forecasts is assumed to vary with time.

The combination weights represent the relative quality, and in order to keep them up to date, only the most recent observations are employed for weight estimation.

Does the extremity of α_0 influence the performance of the combination techniques? The answer is in the affirmative, if we neglect the covariance matrices $\mathbf{W}$ generating a close-to-diagonal optimal weight $\mathbf{L}_0$. These matrices have been already dealt with. For the other choices we can state the tendency that the dominance of $\widehat{\text{SCAT}}$ over *all* other combination techniques is distinct if $|\alpha_0 - 1/2|$ is large. For the very extreme α_0-values ($\mathbf{W}_9$ and $\mathbf{W}_{10}$) the $\widehat{\ }$-variants of WCAT, MCAT and SCAT perform much better than the $\widetilde{\ }$-variants. For the more moderate α_0-values ($\mathbf{W}_1, \ldots, \mathbf{W}_6$) all combined estimators outperform both individual estimators, whereas only one individual estimator is outperformed by the arithmetic mean and the $\widetilde{\ }$-variants for the more extreme α_0-values ($\mathbf{W}_7, \ldots, \mathbf{W}_{10}$).

Does good estimation of the optimal weight $\mathbf{L}_0$ imply good parameter estimation? From comparing Table 3 to Tables 1 and 2 it may be concluded that the quality of weight estimation can only serve as a rough indicator for the quality of parameter estimation: If two methods have errors with respect to $\mathbf{L}_0$ which are *very* different, then these methods will perform differently for parameter estimation as well. But if the differences in weight estimation are relatively small one cannot say which method will be better for parameter estimation. Frequently, from considering the weight errors we would expect one technique to be much better than another, but considering the parameter errors reveals only a small difference. Or it may even happen that weight errors and parameter errors result in a completely different ranking of combination techniques. This is the case for ($\mathbf{W}_2$, $n = 10$) where $\widetilde{\alpha}_0\mathbf{I}_2$ is the best estimator for $\mathbf{L}_0$ but $\mathbf{T}_{\widetilde{\alpha}_0}$ is the worst parameter estimator.

What else can be said about weight estimation? In the last paragraph we stated a rough relationship between the quality of weight estimators and the quality of parameter estimators. Consequently, for weight estimation we can observe similar effects as have been listed above for parameter estimation, e.g. concerning the results for the $\widehat{\ }$- and $\widetilde{\ }$-variants, etc.

Does the quality of estimating $\mathbf{W}$ influence the quality of estimating $\mathbf{L}_0$? The goodness of estimating $\mathbf{W}$ improves with increasing sample size n. The same is true for the estimation of $\mathbf{L}_0$ by $\widehat{\mathbf{L}}_0$. The estimation of $\mathbf{L}_0$ by any of $\widehat{\alpha}_0\mathbf{I}_2$, $\widetilde{\alpha}_0\mathbf{I}_2$, $\widehat{\mathbf{D}}_0$, $\widetilde{\mathbf{D}}_0$ or $\widetilde{\mathbf{L}}_0$ does not necessarily improve as well, since these estimates may converge to wrong combination weights (cf. the paragraph on the sample size n).

Can some covariance matrices $\mathbf{W}$ be better estimated than others? It can be observed that the two best results in the estimation of $\mathbf{W}$ occur for $\mathbf{W}_{10}$ and $\mathbf{W}_7$. These are the only covariance matrices with positive entries only. The corresponding

$\widehat{\text{SCAT}}$ combined estimators $\mathbf{T}_{\widehat{\mathbf{L}}_0}$, however, do not show extraordinarily good results.

5 Conclusions

All covariance adjustment techniques performed well in the above simulation study. In general they outperformed the arithmetic mean $\mathbf{T}_{AM}$ and, thus, they should be preferable. Some comments on the usefulness of the arithmetic mean as a combination method in practical applications follow later in this section.
As a consequence of our results we suggest the following *pre-estimate combination method*:

- Estimate the covariance matrix $\mathbf{W}$
- Calculate the estimated optimal weight $\widehat{\mathbf{L}}_0$
- If $\widehat{\mathbf{L}}_0 \approx \alpha \mathbf{I}$ then combine via $\widehat{\text{WCAT}}$
- Otherwise, if $\widehat{\mathbf{L}}_0 \approx \text{diag}$ then combine via $\widehat{\text{MCAT}}$
- Otherwise combine via $\widehat{\text{SCAT}}$

At least for small sample sizes like $n = 10$ $\widetilde{\text{WCAT}}$ or $\widetilde{\text{MCAT}}$ may be considered as an alternative for $\widehat{\text{WCAT}}$ or $\widehat{\text{MCAT}}$, respectively. If the estimate $\widehat{\alpha}_0$ of the WCAT optimal combination parameter is near to $1/2$ and n is small we may consider $\widetilde{\text{SCAT}}$ as an alternative for $\widehat{\text{SCAT}}$. If the sample size is as large as $n = 50$ we should employ $\widehat{\text{SCAT}}$ throughout.
The results from the simulation study suggest that employing the pre-estimate combination method should be leading to more accurate estimates than using any fixed covariance adjustment technique, the arithmetic mean or an individual estimator.
For the general case where $\mathbf{L}_0$ is not close to a diagonal matrix the possible profit of employing $\widehat{\text{SCAT}}$ is enormous, whereas the possible loss is only small. Furthermore only $\widehat{\text{SCAT}}$ will produce asymptotically optimal results (for large n). The conjectured advantage of the medium and weak variants, based on the fact that fewer parameters have to be estimated, could not be confirmed unless $\mathbf{L}_0$ is approximately diagonal. Regarding this dominance of $\widehat{\text{SCAT}}$ in the general case leads us to the recommendation that additional variables should be used whenever possible. If estimators for a set of parameters are to be combined all parameters should be treated simultaneously by $\widehat{\text{SCAT}}$ instead of treating all the parameters separately.

Based on the findings of Makridakis and Winkler (1983), we conjectured that the use of covariances between the parameter estimators may not be beneficial. This conjecture could not be confirmed in general. Especially for SCAT the $\hat{}$-variant (using the covariances) performed much better than the $\tilde{}$-variant (neglecting the covariances). For MCAT, however, which treats all the components like in the univariate case, and small sample size n we found some evidence that neglecting covariances may be a good alternative. It is worthwhile noting that this confirms the findings of Makridakis and Winkler (1983) since they are using univariate forecasts and their samples for estimating the combination weights are mostly small. But as already stated we would rather search for appropriate additional variables than following univariate strategies, if possible.

A relationship between the quality of combination weight estimation and the quality of parameter estimation was observed, but it was not so close as might have been assumed.

Recall that these results have been obtained using a simulation study with random data from multivariate normal distributions. It seems advisable to analyze how far the good performance of the CA techniques with estimated weights depends on normality. The losses with respect to the CA techniques with optimal combination weights may vary with the chosen distribution, and so may the gains with respect to the arithmetic mean where no weight estimation is necessary. Hence, the covariance adjustment techniques should be investigated under different multivariate distributions including skew distributions where the data are not distributed symmetrically about their expectation.

Furthermore the parameters of the normal distributions are held constant, i.e. the data used to estimate the combination weights are not subject to any structural change. In empirical applications, however, structural change is present frequently. Consequently, the effects of structural change should be investigated as well. Often this phenomenon is treated by using only the latest observations to estimate the combination weights. Thus, small sample sizes are important in this context.

The parameter estimators used in the present study are unbiased. An interesting question is how biased estimators should be dealt with. Related to this is the following: By the choice of the combination weights it is obvious that the above covariance adjustment techniques are designed for unbiased parameter estimators. Then it would be important to judge the effect of using CAT if this assumption of unbiasedness is violated.

While our results (obtained under the somewhat ideal conditions of the simulation study) suggest that the arithmetic mean is 'out' as a method to combine unbiased estimators of a parameter vector under normality, the situation may be different in case of practical applications, like the combination of forecasts. Here the objective of the forecasts is the realization of a random variable, no information on the distributions of the single forecasts is available and even the relative quality of the forecasters will change with time (structural change). Hence, we have to face many of the above mentioned problems at a time. Consequently, the comparison of combining methods may be less favorable for the covariance adjustment techniques and more favorable for the arithmetic mean. Many empirical studies on the *univariate* combination of forecasts (e.g. Makridakis and Winkler (1983)), including recent results by Klapper (1998) dealing with forecasts for German macro economic variables, indicate that the arithmetic mean is quite a good choice under such circumstances. Whether the extra information included by employing the *multivariate* covariance adjustment techniques will lead to an improvement over the arithmetic mean is not clear and may depend on the data under consideration. A further technical report (Troschke (1999)) will investigate this question for the above mentioned data set of forecasts for German macro economic variables.

There are some further questions connected with the topic of this report that leave room for investigations: In the first place, it would be interesting if the results are similar for higher dimensions k of the parameter vector $\boldsymbol{\theta}$. On the one hand there are additional covariances which may be exploited by the CA techniques, but on the other hand the discrepancy with respect to the number of parameters that have to be estimated for the various CA techniques $(1, k, k^2)$ would be much larger than here $(1, 2, 4)$.

Another point of interest is the effect of the number of estimates used for the CA techniques. It could be conjectured that there is a certain effect of saturation, i.e. the performance could improve with every additional estimator, but once a certain number of estimators is reached the effect of adding more estimators becomes very small. Naturally, this will also depend on whether the additional estimators bring in new information on $\boldsymbol{\theta}$.

Finally, the effect of outliers for the covariance adjustment techniques could be investigated. It might be beneficial then to replace the sample covariance matrix as estimator for $\mathbf{W}$ by some robust estimator, like the *minimum volume ellipsoid* estimator or an S-estimator (cf. Rousseeuw (1985)). Preliminary analysis in this direction showed that the use of robust estimators could not improve the results in the study described above. This could be expected, since the multivariate normal distribution does not tend to produce outliers. But even when the random data from the $\mathcal{N}_4(\mathbf{0}, \mathbf{W})$-distribution were contaminated with data from the $\mathcal{N}_4(\mathbf{0}, 2\mathbf{W})$-

distribution (the contamination rate was about 10%), in general the covariance adjustment techniques performed better employing the sample covariance matrix than employing its robust alternatives. On the other hand this setting tends to produce only so-called radial outliers and the portion of outliers is quite small.

Acknowledgements: I want to thank Jürgen Groß and Götz Trenkler 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

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

Dickinson, J.P. (1988): Company forecast accuracy for exponential smoothing models of earnings-per-share data for financial decision making: A comment, *Decision Sciences* **19**, 233-235.

Horn, R.A. and Johnson, C.R. (1985): Matrix Analysis. *Cambridge University Press, Cambridge.*

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

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

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

Makridakis, S. and Winkler, R.L. (1983): Averages of forecasts: Some empirical results, *Management Science* **29**, 987-996.

Odell, P.L., Dorsett, D., Young, D. and Igwe, J. (1989): Estimator models for combining vector estimators, *Mathematical and Computer Modelling*, **12**, 1627-1642.

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.

Rousseeuw, P.J. (1985): Multivariate estimation with high breakdown point, in: Grossmann, W., Pflug, G., Vincze, I. and Wertz, W. (eds.), Mathematical Statistics and Applications, **8**, *Reidel, Dordrecht*, 283-297.

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. (1999): An Extensive Analysis of German Macroeconomic Forecast Data Using Covariance Adjustment Techniques, *Technical Report, Sonderforschungsbereich 475, University of Dortmund* (to appear).