A Selective Procedure For Combining Forecasts

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

Abstract: If there are various forecasts for the same random variable, it is common practice to combine these forecasts in order to obtain a better forecast. But an important question is *how* to perform the combination, especially if the system under investigation is subject to structural changes and, consequently, the best combination method is not the same all of the time. This paper presents a data driven approach, which (for each point of time) selects a combination technique from a given set of combination techniques. Properties and limitations of this selection procedure are investigated using simulated data from normal distributions.

Keywords: Combination of forecasts, selection predictor.

AMS 1991 Subject Classification: 62M20, 62F10

1 Introduction

Let $E_1, \ldots, E_p$ be predictors of a one-dimensional random variable θ. The predictors may or may not be unbiased for θ. Suppose that they can be calculated from data available at each point of time $t \in I\!N$. We assume in this paper that neither of the predictors is best all of the time, but that different predictors may be best at different times. The notion 'best' may refer to any criterion chosen to judge the quality of a predictor such as root mean square error (RMSE), mean absolute deviation (MAD), and so on.
Wondering which of $E_1, \ldots, E_p$ to use for the prediction of θ, one may determine the best predictor from theoretical considerations or from past data and use this predictor in the future. But each predictor may use information that the others neglect. Thus, the idea is to combine the forecasts in order to obtain a better forecast for θ. But how can this be done most efficiently?

Note that a special case of the problem described above is combining *estimators* for a *constant* θ.
There has been a great number of articles in the forecasting literature dealing with the various ways to obtain a good combination, after the idea of combining forecasts was introduced by Bates and Granger (1969). In fact, the number of offered alternatives is so large that one may easily lose sight, not knowing which to prefer. The idea in this article is to preselect a set $\mathcal{S}$ of good combination techniques and let past data decide which of these techniques should be used.
Applying this idea to single predictors means to select the best predictor from a given set of predictors. As indicated above this method is not very promising and can be outperformed by combined forecasts, since they use more information.
But applying the idea to combined forecasts is a different issue, because each combined predictor already contains all the available information. Thus it is interesting to see how this idea works out in practical situations. For this purpose we will conduct a simulation study using normally distributed data. Another analysis using German macroeconomic forecast data will throw additional light on this topic. It will be reported in a future Technical Report (Troschke (1998)).
A different view on the selection predictor is provided by the following thoughts: To choose that combination technique, which was best in the past is a very intuitive way of deciding for one of the many possible combination techniques. The selection predictor makes this decision at each point of time based on the past data available and so may a person who is in charge of deciding for a combination technique. Our analysis judges the effects of this kind of decision making.
Section 2 of this report will introduce the so called *selection predictor*, which formalizes the above idea. Section 3 presents the design of a simulation study conducted to reveal the properties of the selection predictor. Comparison of different predictors will be done using the (empirical) root mean square error criterion, which will be described in Section 4. Section 5 reports and evaluates the results of the simulation study, before Section 6 concludes this paper with some final remarks.

2 The Selection Predictor

If there are several possible methods to forecast the values of a random variable θ we wish to select the most efficient method at each time t. A way to make this selection on the basis of past data is provided by the following definition. The amount of past data used is determined by the choice of the parameter h.
As indicated in the introduction the set $\mathcal{S}$ of possible methods should consist of

predictors using all the available information. In general, these will be combined predictors calculated from a number of single predictors available to the statistician. Hence, the predictors in $\mathcal{S}$ will differ in the way the combination is performed. This includes weighted combinations which differ in the way the combination weights are calculated.

Definition 2.1 (Selection predictor)
Let $\mathcal{S} = \{C_1, \ldots, C_k\}$ be a set of predictors for a one-dimensional random variable θ and let $C_j(i)$ denote the forecast provided by C_j at time i. Furthermore, let $\text{RMSE}(C_1, \theta, t, h), \ldots, \text{RMSE}(C_k, \theta, t, h)$ be the respective root mean square errors of the predictors from $\mathcal{S}$ with respect to θ calculated at time t from the last h points of time, i.e.

$$\text{RMSE}(C_j, \theta, t, h) = \left(\frac{1}{h} \sum_{i=t-h}^{t-1} (C_j(i) - \theta(i))^2 \right)^{1/2} , \quad j = 1, \ldots, k \, .$$

Then the *selection predictor* $S(t,h) = S(C_1(t), \ldots, C_k(t), h)$ at time t on the basis of the past h points of time is defined by the following procedure:

- Identify the predictor $C_{j0} \in \mathcal{S}$ producing the smallest value $\text{RMSE}(C_j, \theta, t, h)$, $j = 1, \ldots, k$.
- The selection predictor at time t is $S(t,h) = C_{j0}(t)$.

The question how h should be chosen will always have to be answered with respect to the system under consideration. One would expect that larger values for h are appropriate if the system exhibits a certain stability in the sense that the relative quality of the predictors does not change too fast. Smaller values, even as small as $h = 1$, should be chosen if this kind of stability is absent. How h should be chosen exactly will depend on the grade of stability in the system.
In forecasting we will often (if not always) observe that the relative quality of the forecasts changes with time. If there are, e.g., two forecasters of economic variables, one forecaster may provide relatively better forecasts when the economy is in a downswing than if the economy is in an upswing. Such situations are often referred to as *structural changes*.
The simplest way to react to such structural changes is to update the combination weights whenever a new combination is to be performed. Bates and Granger (1969) suggest several simple time-varying weights. Diebold and Pauly (1987) propose more sophisticated techniques, all extensions of the standard regression-based

theory of forecast combination. Deutsch, Granger and Teräsvirta (1994) investigate several variants of switching regression models: Here different regression models are employed, dependent on which state the system under investigation is supposed to be in.

Again, the reader is left with a great variety of combination techniques with no guide when to apply which technique. The selective procedure introduced in this paper tries to overcome this difficulty. The set $\mathcal{S}$ may contain combined predictors with simple time-varying weights as well as predictors based on sophisticated regression models or switching regression models. It may also contain combined predictors based on rank techniques (Russel and Adam (1987), Klapper (1998)) or predictors employing covariance adjustment techniques, if more than one variable is to be forecasted (Rao (1966, 1967), Ihorst (1993), Trenkler and Ihorst (1995)). The data driven selection procedure will choose the method which produced the best results in the past and apply this technique. Thus one may possibly benefit from all the suggested models.

In the next sections we will analyse the properties of the selection predictor by means of a simulation study.

3 Design of the Simulation Study

Our simulation study is designed to handle a situation where the relative quality of two predictors varies with time. The study comprises n points of time $t = 1, \ldots, n$. The number n will be referred to as the *length of the study*. Let $X(t)$ and $Y(t)$ be two normally distributed random variables with a common mean $\mu(t)$ but with possibly different variances $\sigma_X^2(t)$ and $\sigma_Y^2(t)$, i.e. $X(t) \sim \mathcal{N}(\mu(t), \sigma_X^2(t))$ and $Y(t) \sim \mathcal{N}(\mu(t), \sigma_Y^2(t))$. While $\mu \equiv 0$ is held fixed, the relation between $\sigma_X^2(t)$ and $\sigma_Y^2(t)$ changes during the study as will be described below.

Since μ is assumed to be constant we are in the special case of estimation mentioned in the introduction. The goal is to estimate μ in such a way that the associated root mean square error (RMSE) is minimized. For this purpose at each point of time $t \in \{1, \ldots, n\}$ in the study we observe independent samples $X_1(t), \ldots, X_{10}(t)$ and $Y_1(t), \ldots, Y_{10}(t)$ from the respective random variables $X(t)$ and $Y(t)$.

Two standard estimators of μ are the respective means of the X- and Y-samples, i.e.

$$\overline{X}(t) = \frac{1}{10}\sum_{i=1}^{10} X_i(t) \qquad \text{and} \qquad \overline{Y}(t) = \frac{1}{10}\sum_{i=1}^{10} Y_i(t) \quad t = 1, \ldots, n\ .$$

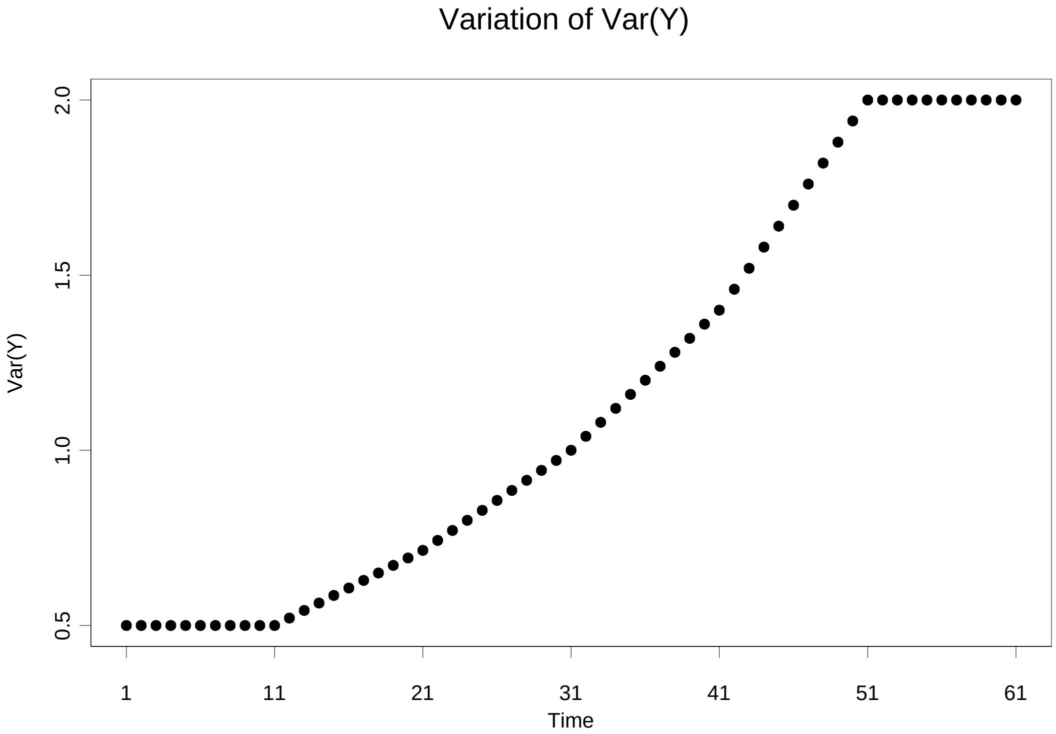

Figure 1: Variation of σ_Y^2 during the study for study length $n = 61$.

Each of these two estimators is based on only one of the two samples and, thus, does not contain all the available information. Consequently, combined estimators should be used. Combined estimators calculated from the two standard estimators include convex combinations $T_\alpha(t)$ of $\overline{X}(t)$ and $\overline{Y}(t)$, i.e.

$$T_\alpha(t) = \alpha\overline{X}(t) + (1-\alpha)\overline{Y}(t) \ , \alpha \in I\!R, \ t = 1, \ldots, n \ .$$

A look at the way we intend to change σ_X^2 and σ_Y^2 helps us determining the combined estimators we should consider: While holding $\sigma_X^2 = 1$ fixed we will vary σ_Y^2 in the following way: $\sigma_Y^2 = 1/2$ is held constant during the first sixth of the study, then it increases linearly to $\sigma_Y^2 = 5/7$ during the second sixth of the study, then linear increases to $\sigma_Y^2 = 1$, $\sigma_Y^2 = 7/5$ and $\sigma_Y^2 = 2$ follow in the subsequent sixths of the study. In the final sixth $\sigma_Y^2 = 2$ is held constant again. Obviously, we can state that the smaller we choose n (the length of the study), the faster we vary σ_Y^2. The way how σ_Y^2 is varied can be seen from Figure 1 for $n = 61$.
Since $\overline{X}(t)$ and $\overline{Y}(t)$ are unbiased for all $t = 1, \ldots, n$ the consequence of changing σ_Y^2 as described above is, that at the beginning of the study $\overline{Y}(t)$ is a better estimator

(in terms of root mean square error) than $\overline{X}(t)$, in the middle of the study they are equally good and at the end $\overline{X}(t)$ is better than $\overline{Y}(t)$. By this we create a process in which the best (combined) estimator will not be the same all the time, a situation in which the use of selection estimators may be beneficial.

Varying σ_Y^2 as described above includes only one change from $\sigma_Y^2 < \sigma_X^2$ to $\sigma_Y^2 > \sigma_X^2$. Other variation schemes could be applied as well, especially such schemes where the ranking of σ_Y^2 and σ_X^2 changes more frequently during the study. We believe, however, that the effect of such schemes can be approximated by regarding the variation described above with small values for n.

The respective values 1/2, 5/7, 1, 7/5 and 2 for σ_Y^2 correspond to the optimal α-values 1/3, 5/12, 1/2, 7/12 and 2/3 in the convex combination $T_\alpha(t) = \alpha\overline{X}(t) + (1-\alpha)\overline{Y}(t)$. Since $\overline{X}(t)$ and $\overline{Y}(t)$ are independent the optimal value α_0 is given by (cf. Bates and Granger (1969))

$$\alpha_0 = \frac{\mathrm{Var}(\overline{Y}(t))}{\mathrm{Var}(\overline{X}(t)) + \mathrm{Var}(\overline{Y}(t))} = \frac{\sigma_Y^2(t)}{\sigma_X^2(t) + \sigma_Y^2(t)} \; .$$

Thus, $T_{1/3}(t)$ is optimal at the beginning of the study and $T_{2/3}(t)$ is optimal at the end. Hence, we will use $\mathcal{S} = \{T_{1/3}, T_{2/3}\}$ as the set $\mathcal{S}$ from which the selection estimator will choose, i.e. the selection estimator will choose from

$$T_{1/3}(t) = \frac{1}{3}\overline{X}(t) + \frac{2}{3}\overline{Y}(t) \qquad \text{and} \qquad T_{2/3}(t) = \frac{2}{3}\overline{X}(t) + \frac{1}{3}\overline{Y}(t) \; .$$

We deliberately chose this very simple constellation of estimators with constant weights. We believe that the basic properties of the selection estimator can be seen best using this choice, since they are not masked by difficulties that may arise when using more sophisticated combination techniques. If we used, for example, weighted means of $\overline{X}$ and $\overline{Y}$ with weights dependent on past data, the estimation of the weights would be an additional problem disguising the properties of the selection estimator.

With $\mathcal{S} = \{T_{1/3}, T_{2/3}\}$ the selection estimator can only choose between $T_{1/3}$ and $T_{2/3}$. Nevertheless, the selection estimator may well be better than both, $T_{1/3}$ *and* $T_{2/3}$, since it may choose the better combined estimator *at each point of time* during the study.

To calculate the selection estimator we must be aware that it relies on the performance of the estimators from $\mathcal{S}$ in the past. That is why a certain amount of past data is needed before the selection procedure may be employed. These data are provided by the first ten points of time (Phase I).

Consequently, Phase II where the performances of all estimators under consideration are compared, starts with $t = 11$ and ends with $t = n$. The comparison is done in terms of (empirical) root mean square errors, see Section 4 for details.
In the simulation study the whole process (Phase I and II) is repeated 1 000 times. The results from the single repetitions are recorded for further evaluation with the aim to answer the following questions connected with the selection estimator:

(1) How does the selection estimator perform relative to the single estimators $\overline{X}$ and $\overline{Y}$?

(2) How does the selection estimator perform relative to the arithmetic mean of $\overline{X}$ and $\overline{Y}$?

(3) How does the selection estimator perform relative to the estimators $T_{1/3}$ and $T_{2/3}$ from which it selects?

(4) How is the quality of the selection estimator affected by the speed of the change of σ_Y^2?

(5) How much past data should the selection procedure use to make its choice from the set of estimators $\mathcal{S}$? (This refers to the choice of the parameter h in the definition of the selection estimator.)

(6) What is the effect if we enlarge the set $\mathcal{S}$ of estimators from which the selection procedure chooses?

The first five questions will be answered by a simulation study, in which we will calculate the (Phase II) RMSE-values of $\overline{X}$, $\overline{Y}$, $T_{1/3}$, $T_{2/3}$, selection estimators from $\mathcal{S} = \{T_{1/3}, T_{2/3}\}$, and of the arithmetic mean

$$T_{1/2} = \frac{1}{2}\overline{X} + \frac{1}{2}\overline{Y}$$

of $\overline{X}$ and $\overline{Y}$, respectively. Note that $T_{1/2}$ is not only the arithmetic mean of $\overline{X}$ and $\overline{Y}$ but also the arithmetic mean of $T_{1/3}$ and $T_{2/3}$ and of $T_{5/12}$ and $T_{7/12}$, which will be introduced later on.
Question 4 will be answered by looking at RMSE-values for varying lengths n of the simulation study. σ_Y^2 steadily increases from $1/2$ to 2 regardless of the choice for n. Consequently, if n is a small value the change of σ_Y^2 is quite fast, whereas for large values of n the change is very slow. In our simulation study we will consider $n = 19$, $n = 31$, $n = 61$, $n = 121$, $n = 181$ and $n = 241$.

To investigate Question 5 within the first study we will employ two different strategies for the calculation of the selection estimator. On the one hand at time t the selection procedure will choose the estimator with the smallest RMSE-value determined from the past 10 points of time, i.e. we choose $h = 10$ in the definition of the selection estimator. On the other hand at time t the selection procedure will choose the estimator with the smallest RMSE-value determined from all points of time up to $t-1$, i.e. we choose $h = t-1$. The latter strategy is often employed with the idea not to waste any information from past data, while the former presumes that past data may become too old to be valid for the current point of time.

Of course, other choices for h may be reasonable. As indicated in Section 2 the best choice for h will always depend on the properties of the system under investigation. In order to throw additional light on this topic a second simulation study is carried out. Here the RMSE-values of the selection estimator from $\mathcal{S} = \{T_{1/3}, T_{2/3}\}$ are calculated for varying h. For $n = 19$ and $n = 31$ we will investigate $h = 1$, $h = 2$, $h = 3$, $h = 5$, $h = 7$ and $h = 10$, for $n = 61$, $n = 121$, $n = 181$ and $n = 241$ we will additionally consider $h = 15$ and $h = 20$. Since we want to use $h = 15$ and $h = 20$ we prolonged Phase I for the larger study lengths n, i.e. for $n = 61$, $n = 121$, $n = 181$ and $n = 241$ we will have Phase I from $t = 1$ to $t = 20$ and Phase II from $t = 21$ to $t = n$.

The final Question (6) will be answered by a third simulation study. Here we will include further estimators in the set $\mathcal{S}$ from which the selection estimator chooses. The further estimators will be the arithmetic mean $T_{1/2}$,

$$T_{5/12} = \frac{5}{12}\overline{X} + \frac{7}{12}\overline{Y} \qquad \text{and} \qquad T_{7/12} = \frac{7}{12}\overline{X} + \frac{5}{12}\overline{Y}\,.$$

These estimators are also optimal at some time during the study, $T_{5/12}$ after one third, $T_{7/12}$ after two thirds, and $T_{1/2}$ after half of the study.

It should be noted that the three simulation studies are independent of each other, i.e. they are not using the same random data but new data are generated for each study. Consequently, the corresponding tables will not show exactly the same RMSE-values for the same choices of parameters. By comparing these values the reader may get an impression of the variation of the average from 1 000 simulation runs.

4 Error Measurement

To judge the performance of an estimator T, we will calculate the (empirical) RMSE of T with respect to μ (here $\mu \equiv 0$) on the basis of the data from $t = 11$ to $t = n$

(Phase II), i.e.

$$\mathrm{RMSE}(T, \mu, n+1, n-10) = \left(\frac{1}{n-10} \sum_{t=11}^{n} (T(t) - \mu)^2 \right)^{1/2} .$$

Here $T(t)$ denotes the estimate provided by T at time t.
The arithmetic mean $T_{1/2} = (1/2)(\overline{X} + \overline{Y})$ of the single estimators under consideration is a simple but very efficient combined estimator. It proves successful in many practical studies and very often outperforms much more sophisticated combination techniques. A very comprehensive study of this kind is reported by Makridakis et al. (1982), Makridakis and Winkler (1983), and Winkler and Makridakis (1983).
The arithmetic mean is more robust with respect to changes of the relative quality of the estimators than most other combination techniques. This appears to be the reason for its success and this is why the performance of the arithmetic mean is the touchstone against which all other combination techniques have to be measured. Consequently, throughout this paper we will give our results not in terms of the respective root mean square errors (i.e. $\mathrm{RMSE}(T, \mu)$) but in terms of the root mean square errors relative to the root mean square error of the arithmetic mean (i.e. $\mathrm{RMSE}(T, \mu)/\mathrm{RMSE}(T_{1/2}, \mu)$). These values will be referred to as 'relative RMSE-values'.

5 Results

Table 1 shows the RMSE-values (relative to the RMSE of the arithmetic mean) of $\overline{X}$, $\overline{Y}$, $T_{1/3}$, $T_{2/3}$, and of two selection estimators calculated from $\mathcal{S} = \{T_{1/3}, T_{2/3}\}$. The first selection estimator uses past data from the last 10 points of time ($h = 10$), while the second uses all available past data ($h = t - 1$). All values are average values from 1 000 simulation runs as described in Section 3. The values have been truncated after the fourth decimal.

(1) How does the selection estimator perform relative to the the single estimators $\overline{X}$ and $\overline{Y}$?

As can be seen from Table 1 the selection estimators outperform $\overline{X}$ and $\overline{Y}$ by far (with the only exception being the case $n = 19$ where the selection estimator using all points of time is slightly worse than $\overline{X}$). This had to be expected, since the selection estimators choose from combined estimators and, thus, use more information than $\overline{X}$ or $\overline{Y}$ separately.
The above impression is confirmed by Table 4 in the appendix. Here we have recorded how often (in percent of the 1 000 simulation runs) the selection estimator was better

n	$\overline{X}$	$\overline{Y}$	$T_{1/3}$	$T_{2/3}$	$S(T_{1/3}, T_{2/3})$ 10 p.o.t.	all p.o.t.
19	1.0542	1.7628	1.2065	0.8887	1.0085	1.0634
31	1.1430	1.6695	1.1670	0.9325	0.9718	1.0421
61	1.1822	1.6213	1.1478	0.9525	0.9354	1.0127
121	1.2108	1.5991	1.1375	0.9646	0.9288	0.9945
181	1.2154	1.5899	1.1341	0.9676	0.9233	0.9882
241	1.2237	1.5828	1.1310	0.9712	0.9245	0.9867

Table 1: Selection from two estimators: Average RMSE-values (relative to the RMSE of the arithmetic mean) from 1 000 simulation runs.

than both, $\overline{X}$ and $\overline{Y}$ / one of them / none of them. The selection estimators were never worse than both, and from $n = 31$ on they were better than both in most of the simulation runs. Again, the selection estimator using 10 points of time showed better results than the selection estimator using all points of time.

(2) How does the selection estimator perform relative to the arithmetic mean of $\overline{X}$ and $\overline{Y}$?

From Table 1 we see that the selection estimator using 10 points of time outperforms the arithmetic mean from $n = 31$ on, by up to 7.5 percent. For $n = 19$ these two estimators are approximately equal. The selection estimator using all points of time is worse than $T_{1/2}$ in general. Only from $n = 121$ on the selection estimator wins by a small margin.

Table 5 in the appendix reveals the distribution of the relative RMSE-values of the estimators. To see how often (in percent of the 1 000 simulation runs) the estimators could outperform the arithmetic mean we may have a look at the third column of Table 5. Columns four through nine allow to judge by what margin the estimators were better / worse than $T_{1/2}$.

The three histograms in Figure 2 correspond to specific rows of Table 5.

For the short study length $n = 19$ we see that the selection estimator from $\mathcal{S} = \{T_{1/3}, T_{2/3}\}$ using 10 points of time is about as good as $T_{1/2}$: The distribution of the relative RMSE-values is almost symmetric around 1. For larger study lengths n (e.g. $n = 121$) we observe that the distribution changes in favour of the selection estimator. A gain of 5 to 10 percent with respect to $T_{1/2}$ is observed most frequently. The performance of the selection estimator from $\mathcal{S} = \{T_{1/3}, T_{2/3}\}$ using all points of time is much worse.

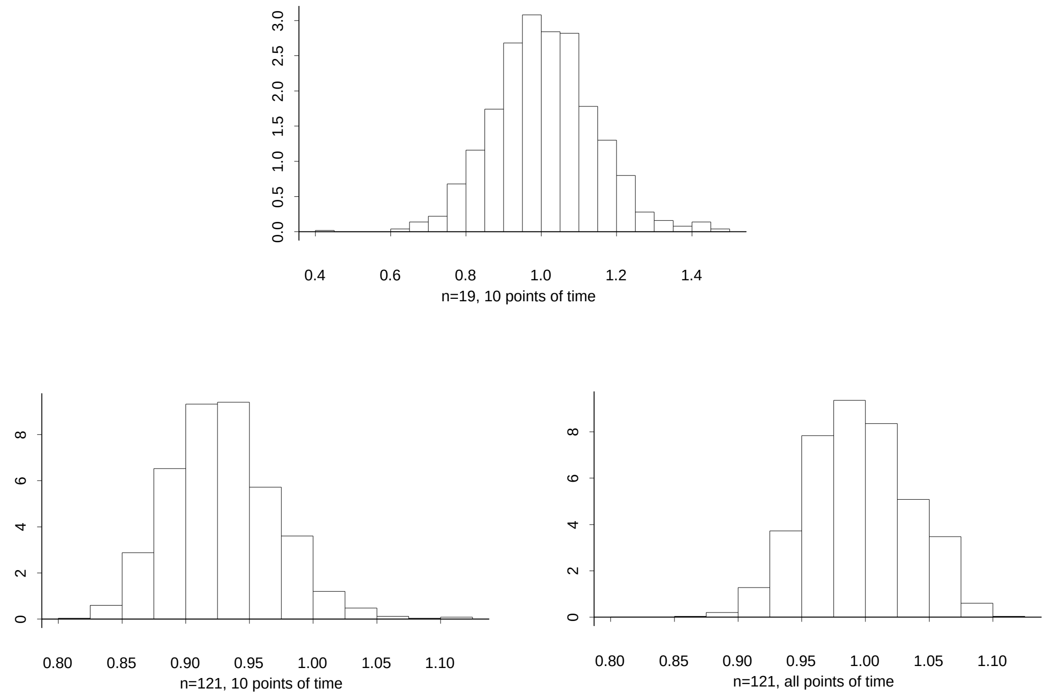

Figure 2: Distributions of the relative RMSE-values for selection estimators from $\mathcal{S} = \{T_{1/3}, T_{2/3}\}$.

(3) How does the selection estimator perform relative to the estimators $T_{1/3}$ and $T_{2/3}$ from which it selects?

Of course one would wish that the selection estimator outperforms the estimators from which it selects or that it is as good as the best estimator from $\mathcal{S}$. But since we do not know beforehand which of the estimators from $\mathcal{S}$ will be the best, we might also be satisfied if the selection estimator outperforms many of the estimators in $\mathcal{S}$. For $\mathcal{S} = \{T_{1/3}, T_{2/3}\}$ we can see from Table 1 that the selection estimator using ten points of time is better than $T_{1/3}$ for all study lengths n and better than $T_{2/3}$ from $n = 61$ on. The selection estimator using all points of time is better than $T_{1/3}$ for all n as well, but it is also worse than $T_{2/3}$ for all n.

Table 6 in the appendix records how often (in percent of the 1 000 simulation runs) the selection estimator was better than both, $T_{1/3}$ and $T_{2/3}$ / one of them / none of them. For the smaller study lengths $n = 19$ and $n = 31$ the selection estimator using 10 points of time is at least as good as the better estimator of $T_{1/3}$ and $T_{2/3}$ in only 15 to 18 percent of the simulation runs. But it is better than at least one of these estimators in 86 to 92 percent of the simulation runs. For larger n these percentages rise enormously: The selection estimator using 10 points of time outperforms $T_{1/3}$ *and* $T_{2/3}$ in most of the simulation runs. The selection estimator using all points of time performs much worse again.

(4) How is the quality of the selection estimator affected by the speed of the change of σ_Y^2?

Recall that the speed of the change of σ_Y^2 is manipulated by varying the study length n. As indicated by the answers to the previous questions, it must be stated that the performance of the selection estimators becomes the better the larger the study length n grows, i.e. the slower σ_Y^2 changes. Since there is not much change in the relative RMSE-values after $n = 121$ we can assume, however, that a state of saturation is reached (Table 1).

For the smallest n in our study ($n = 19$) the selection estimator using 10 points of time is about as good a choice as the arithmetic mean and for larger n-values this selection estimator outperforms the arithmetic mean and exhibits satisfying properties.

(5) How much past data should the selection procedure use to make its choice from the set of estimators $\mathcal{S}$?

Regarding all questions considered above the selection estimator using ten points of time performed better than the selection estimator using all points of time. This is confirmed by a look at Table 7 in the appendix, which gives head-to-head results for these two estimators.

	$h=1$	$h=2$	$h=3$	$h=5$	$h=7$	$h=10$	$h=15$	$h=20$
$n=19$	1.0030	0.9826	0.9778	0.9721	0.9835	1.0100	—	—
$n=31$	1.0112	0.9876	0.9768	0.9661	0.9634	0.9686	—	—
$n=61$	1.0115	0.9875	0.9721	0.9563	0.9494	0.9431	0.9438	0.9496
$n=121$	1.0014	0.9808	0.9646	0.9461	0.9364	0.9276	0.9231	0.9211
$n=181$	1.0067	0.9812	0.9655	0.9457	0.9354	0.9265	0.9203	0.9181
$n=241$	1.0047	0.9788	0.9643	0.9444	0.9330	0.9245	0.9172	0.9148

Table 2: Selection estimators from $\mathcal{S} = \{T_{1/3}, T_{2/3}\}$: Average RMSE-values (relative to the RMSE of the arithmetic mean) from 1000 simulation runs

Even for $n = 19$ we see that *10 points of time* is at least as good as *all points of time* in 89 percent of the simulation runs, and for larger n to choose *10 points of time* is better in 75 to 99 percent of the simulation runs. Hence, we can give the advice not to use too old data for the selection process: The farther past is not so important compared to the latest performances of the estimators in $\mathcal{S}$. Data from how far back should be considered will depend on the special situation under investigation.

The second simulation study, reported in Table 2, allows some additional insight concerning this topic. With the usual choices for n we investigated several choices for h and observed the relative RMSE-values (averages from 1000 simulation runs, all values truncated after the fourth decimal). Recall that the relative RMSE-values for $n = 61$, $n = 121$, $n = 181$ and $n = 241$ have been calculated on the basis of the data from $t = 21$ to $t = n$, while the relative RMSE-values for $n = 19$ and $n = 31$ have been calculated using the data from $t = 11$ to $t = n$.

We observe that the slower we change σ_Y^2, i.e. the larger we choose the study length n, the larger we should choose the parameter h, which represents the amount of past data used in the selection procedure.

For fixed n we see that the relative RMSE-values are high for small h. With increasing h the RMSE-values go down first, but from some h on they go up again. This indicates that there is an optimal choice for h dependent on the choice of n. This confirms the intuition that one should neither use too few data (i.e. neglect information) nor too old data.

Reading Table 2 columnwise we observe that the performance of the selection estimator with smaller choices for h is almost independent of the speed of the change in σ_Y^2. For larger h we can confirm that there is a certain point of saturation regarding the study length n.

All in all we see that the selection estimators generally outperformed the arithmetic mean whenever the parameter h was reasonably chosen.

(6) What is the effect if we enlarge the set $\mathcal{S}$ of estimators from which the selection procedure chooses?

Table 3 collects the RMSE-values (relative to the RMSE of the arithmetic mean) of several selection estimators based on different sets $\mathcal{S}$ with up to seven estimators. As a consequence from the results of the first two studies, we chose the data on which the selection procedure is based to comprise the last ten points of time only (and Phase II starts with $t = 11$). Also, it seems to be sufficient to investigate $n = 19$, $n = 31$, $n = 61$ and $n = 121$. Despite the recommendation given in Sections 1 and 2, we have also considered sets $\mathcal{S}$ including the non combined estimators $\overline{X}$ and $\overline{Y}$. The conjecture that the use of these estimators is not beneficial is confirmed by the outcome of the simulation. Again, all values are average values from 1 000 simulation runs which have been truncated after the fourth decimal.

For the cases where $\mathcal{S}$ consists of only two estimators we get similar results concerning the performance of the selection estimators relative to the estimators from $\mathcal{S}$ as in Question 3 above. Of course, for $\mathcal{S} = \{\overline{X}, \overline{Y}\}$ the relative RMSE-values are far higher than for the sets $\mathcal{S}$ consisting of combined estimators.

Whenever a set $\mathcal{S}$ consisting of combined estimators only, is enhanced by $\overline{X}$ and $\overline{Y}$ the relative RMSE-values drop by about 0.03. Hence, we can reassure that $\mathcal{S}$ should consist of combined estimators only.

If we restrict our considerations to sets $\mathcal{S}$ with combined estimators only, we can state that adding a further estimator to $\mathcal{S}$ almost allways reduces the relative RMSE (for $n = 19$ there are minor inconsistencies in that sense). It is obvious, however, that the effect of adding a further estimator is smaller, if the number of estimators in $\mathcal{S}$ is larger. Consequently, one might restrict oneself to employ a set $\mathcal{S}$ with a few good estimators rather than a set with a large number of estimators.

Similar to the third column of Table 5, Table 8 in the appendix shows how often (in percent from the 1 000 simulation runs) the selection estimators were able to outperform the arithmetic mean. While this was the case in about 50 percent of the simulation runs for $n = 19$, the percentage rises up to 98 percent for larger study lengths.

Finally, Table 9 in the appendix records how often (in percent from the 1 000 simulation runs) the selection estimators could outperform the estimators from $\mathcal{S}$. Given are the percentages for the cases when the selection estimator was at least as good as j estimators from $\mathcal{S}$, but not as good as $j + 1$ estimators for $j = 0, \ldots, 5$.

For $n = 19$ the selection estimators are in the midfield for most of the simulation runs, i.e. they are at least as good as half of the estimators from $\mathcal{S}$, but there

$\mathcal{S}$	$n = 19$	$n = 31$	$n = 61$	$n = 121$
$\{\overline{X}\}$	1.0610	1.1169	1.1943	1.2109
$\{\overline{Y}\}$	1.7516	1.6837	1.6167	1.5971
$\{T_{1/3}\}$	1.2017	1.1745	1.1448	1.1370
$\{T_{2/3}\}$	0.8939	0.9228	0.9568	0.9649
$\{T_{5/12}\}$	1.0915	1.0774	1.0616	1.0575
$\{T_{7/12}\}$	0.9323	0.9470	0.9641	0.9682
$\{T_{1/3}, T_{2/3}\}$	1.0144	0.9646	0.9389	0.9258
$\{T_{5/12}, T_{7/12}\}$	0.9928	0.9679	0.9554	0.9490
$\{\overline{X}, \overline{Y}\}$	1.3438	1.2163	1.1503	1.1150
$\{T_{1/3}, T_{1/2}, T_{2/3}\}$	1.0022	0.9564	0.9351	0.9231
$\{T_{5/12}, T_{1/2}, T_{7/12}\}$	0.9924	0.9659	0.9552	0.9487
$\{\overline{X}, T_{1/2}, \overline{Y}\}$	1.0606	1.0323	1.0115	1.0019
$\{T_{1/3}, T_{5/12}, T_{7/12}, T_{2/3}\}$	0.9999	0.9557	0.9341	0.9220
$\{\overline{X}, T_{1/3}, T_{2/3}, \overline{Y}\}$	1.0435	0.9957	0.9652	0.9532
$\{\overline{X}, T_{5/12}, T_{7/12}, \overline{Y}\}$	1.0397	1.0048	0.9812	0.9715
$\{T_{1/3}, T_{5/12}, T_{1/2}, T_{7/12}, T_{2/3}\}$	0.9996	0.9538	0.9340	0.9217
$\{\overline{X}, T_{1/3}, T_{1/2}, T_{2/3}, \overline{Y}\}$	1.0314	0.9876	0.9615	0.9505
$\{\overline{X}, T_{5/12}, T_{1/2}, T_{7/12}, \overline{Y}\}$	1.0393	1.0029	0.9810	0.9711
$\{\overline{X}, T_{1/3}, T_{5/12}, T_{7/12}, T_{2/3}, \overline{Y}\}$	1.0292	0.9870	0.9605	0.9495
$\{\overline{X}, T_{1/3}, T_{5/12}, T_{1/2}, T_{7/12}, T_{2/3}, \overline{Y}\}$	1.0288	0.9851	0.9603	0.9491

Table 3: Selection from up to seven estimators: Average RMSE-values (relative to the RMSE of the arithmetic mean) from 1 000 simulation runs.

is a considerable number of simulation runs (7 to 13 percent) where the selection estimator is even worse than all the estimators from $\mathcal{S}$. For larger study lengths n, and hence slower change of σ_Y^2, the selection estimators can outperform more and more of the estimators from $\mathcal{S}$. For $n = 121$ the selection estimators are better than all of the estimators from $\mathcal{S}$ in about 90 percent of the simulation runs.

6 Conclusions

It has been shown that the selection predictor approach is very promising if the relative quality of the predictors does not change too fast. Hence, this approach might be successful in situations where forecasts are done quite frequently, whereas it might be less beneficial, when there is a rather long period between forecasts. Here, the meaning of 'quite frequently' or of 'a rather long time' must be determined regarding the topic under investigation.

Another fact that should be payed attention to is that the data which form the basis for the selection process should not be too old, so that the selection process is sensitive to changes in the relative quality of the predictors. On the other hand, the amount of past data used in the selection procedure should not be too small, in order not to waste valuable information. We found evidence that there is an optimal compromise between these two demands, i.e. an optimal choice of the parameter h. If the relative quality of the predictors does not change too fast and if h is reasonably chosen a selection predictor may outperform all predictors from the set $\mathcal{S}$ as well as the arithmetic mean of the single forecasts.

Including additional (combined) predictors in $\mathcal{S}$ improves on the performance of the selection predictor, especially for larger values of n, i.e. slower changes of the relative quality of the predictors. Of course, the additional predictors should be reasonably chosen, offering a good alternative to the predictors already available in $\mathcal{S}$ for at least some possible cases. A certain point of saturation could be observed regarding the number of included predictors as well as regarding the slowness of the change of σ_Y^2.

Including non combined predictors in $\mathcal{S}$ reduces the quality of the selection predictor: If a set $\mathcal{S}$ consisting of combined predictors only, is enhanced with noncombined predictors the performance of the selection predictor gets worse.

All the above results have been deduced from a simulation study where the predictors in $\mathcal{S}$ are quite simple, in the sense that they are convex combinations of $\overline{X}$ and $\overline{Y}$ with *fixed* combination weights. Using the selection procedure with combination techniques that need to estimate the weights from past data has an important consequence: The general approach when using such combination techniques is to split the available data into two parts. Then the first part is used to estimate the combination weights and the second part is used to judge the performance of the predictors. When employing selection procedures, however, the available data must be split in *three* parts. To calculate and judge the selection predictor by the third part of the data, we need to know about the performance of the predictors from $\mathcal{S}$ within the second part. But to calculate the predictors from $\mathcal{S}$ for the second part of the data, we need to estimate the respective combination weights from the first part of the data. Thus, employing selection predictors requires an additional splitting of the data: Phases II and I from Section 3 need to be supplemented by a Phase 0.

The selective procedure developed in this report is of interest as a method to choose from a great many of possible combination techniques. The necessity to introduce this procedure arose from an investigation of German macro economic forecast data using covariance adjustment techniques (introduced by Rao (1966, 1967), Ihorst (1993), Trenkler and Ihorst (1995)). This work is presented in another technical report (Troschke (1998)).

References

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

DEUTSCH, M., GRANGER, C.W.J. and TERÄSVIRTA, T. (1994): The combination of forecasts using changing weights, *International Journal of Forecasting*, **10**, 47-57.

DIEBOLD, F.X. and PAULY, P. (1987): Structural change and the combination of forecasts, *Journal of Forecasting*, **6**, 21-40.

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.*

MAKRIDAKIS, S., ANDERSEN, A., CARBONE, R., FILDES, R., HIBON, M., LEWANDOWSKI, R., NEWTON, J., PARZEN, E. and WINKLER, R. (1982): The accuracy of extrapolation (time series) methods: Results of a forecasting competition, *Journal of Forecasting* **1**, 111-153.

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

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.

RUSSELL, T.D. and ADAM JR., E.E. (1987): An empirical evaluation of alternative forecast combinations, *Management Science*, **33**, 1267-1276.

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. (1998): An Extensive Analysis of German Macroeconomic Forecast Data Using Covariance Adjustment Techniques, *Technical Report, Sonderforschungsbereich 475, University of Dortmund* (to appear).

WINKLER, R.L. and MAKRIDAKIS, S. (1983): The combination of forecasts, *Journal of the Royal Statistical Society A* **146**, 150-157.

A Tables

	n	better than both	equal to the better	inbetween	equal to the worse	worse than both
10 p.o.t.	19	0.498	0.000	0.502	0.000	0.000
	31	0.830	0.000	0.170	0.000	0.000
	61	0.993	0.000	0.007	0.000	0.000
	121	1.000	0.000	0.000	0.000	0.000
	181	1.000	0.000	0.000	0.000	0.000
	241	1.000	0.000	0.000	0.000	0.000
all p.o.t.	19	0.416	0.000	0.584	0.000	0.000
	31	0.688	0.000	0.312	0.000	0.000
	61	0.967	0.000	0.033	0.000	0.000
	121	1.000	0.000	0.000	0.000	0.000
	181	1.000	0.000	0.000	0.000	0.000
	241	1.000	0.000	0.000	0.000	0.000

Table 4: Selection estimators from $\mathcal{S} = \{T_{1/3}, T_{2/3}\}$ vs. $\overline{X}$ and $\overline{Y}$.

n	$\mathcal{S}$	Interval for relative RMSE-values						
		$[0,1]$	$[0,0.9)$	$[0.9,0.95)$	$[0.95,1)$	$[1,1.05)$	$[1.05,1.1)$	$[1.1,\infty)$
19	$\{\overline{X}\}$	0.491	0.350	0.063	0.078	0.070	0.064	0.375
	$\{\overline{Y}\}$	0.005	0.002	0.002	0.001	0.005	0.005	0.985
	$\{T_{1/3}\}$	0.032	0.006	0.011	0.015	0.039	0.075	0.854
	$\{T_{2/3}\}$	0.848	0.592	0.151	0.105	0.067	0.044	0.041
	$\{T_{1/3}, T_{2/3}\}$ 10 p.o.t.	0.488	0.200	0.134	0.154	0.142	0.141	0.229
	$\{T_{1/3}, T_{2/3}\}$ all p.o.t.	0.319	0.134	0.087	0.098	0.136	0.142	0.403
31	$\{\overline{X}\}$	0.276	0.142	0.063	0.071	0.081	0.088	0.555
	$\{\overline{Y}\}$	0.000	0.000	0.000	0.000	0.002	0.001	0.997
	$\{T_{1/3}\}$	0.015	0.001	0.000	0.014	0.043	0.132	0.810
	$\{T_{2/3}\}$	0.814	0.344	0.257	0.213	0.117	0.044	0.025
	$\{T_{1/3}, T_{2/3}\}$ 10 p.o.t.	0.653	0.222	0.206	0.225	0.156	0.106	0.085
	$\{T_{1/3}, T_{2/3}\}$ all p.o.t.	0.311	0.057	0.096	0.158	0.201	0.244	0.244
61	$\{\overline{X}\}$	0.081	0.012	0.021	0.048	0.100	0.128	0.691
	$\{\overline{Y}\}$	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	$\{T_{1/3}\}$	0.001	0.000	0.000	0.001	0.023	0.124	0.852
	$\{T_{2/3}\}$	0.832	0.140	0.351	0.341	0.144	0.020	0.004
	$\{T_{1/3}, T_{2/3}\}$ 10 p.o.t.	0.874	0.281	0.339	0.254	0.101	0.022	0.003
	$\{T_{1/3}, T_{2/3}\}$ all p.o.t.	0.404	0.019	0.119	0.266	0.341	0.200	0.055
121	$\{\overline{X}\}$	0.011	0.000	0.002	0.009	0.036	0.102	0.851
	$\{\overline{Y}\}$	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	$\{T_{1/3}\}$	0.000	0.000	0.000	0.000	0.003	0.128	0.869
	$\{T_{2/3}\}$	0.844	0.033	0.319	0.492	0.141	0.014	0.001
	$\{T_{1/3}, T_{2/3}\}$ 10 p.o.t.	0.952	0.251	0.468	0.233	0.042	0.004	0.002
	$\{T_{1/3}, T_{2/3}\}$ all p.o.t.	0.561	0.006	0.125	0.430	0.336	0.102	0.001
181	$\{\overline{X}\}$	0.000	0.000	0.000	0.000	0.021	0.051	0.928
	$\{\overline{Y}\}$	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	$\{T_{1/3}\}$	0.000	0.000	0.000	0.000	0.002	0.109	0.889
	$\{T_{2/3}\}$	0.867	0.006	0.278	0.583	0.130	0.003	0.000
	$\{T_{1/3}, T_{2/3}\}$ 10 p.o.t.	0.986	0.235	0.565	0.186	0.014	0.000	0.000
	$\{T_{1/3}, T_{2/3}\}$ all p.o.t.	0.648	0.000	0.109	0.539	0.324	0.028	0.000
241	$\{\overline{X}\}$	0.000	0.000	0.000	0.000	0.003	0.032	0.965
	$\{\overline{Y}\}$	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	$\{T_{1/3}\}$	0.000	0.000	0.000	0.000	0.000	0.094	0.906
	$\{T_{2/3}\}$	0.877	0.001	0.196	0.680	0.123	0.000	0.000
	$\{T_{1/3}, T_{2/3}\}$ 10 p.o.t.	0.988	0.201	0.615	0.172	0.012	0.000	0.000
	$\{T_{1/3}, T_{2/3}\}$ all p.o.t.	0.689	0.000	0.075	0.614	0.303	0.008	0.000

Table 5: Distribution of RMSE-values (relative to the RMSE of the arithmetic mean) from 1 000 simulation runs for various selection estimators.

	n	better than both	equal to the better	inbetween	equal to the worse	worse than both
10 p.o.t.	19	0.026	0.123	0.715	0.034	0.102
	31	0.175	0.013	0.736	0.000	0.076
	61	0.667	0.000	0.326	0.000	0.007
	121	0.874	0.000	0.124	0.000	0.002
	181	0.951	0.000	0.049	0.000	0.000
	241	0.972	0.000	0.028	0.000	0.000
all p.o.t.	19	0.000	0.127	0.605	0.153	0.115
	31	0.000	0.063	0.728	0.116	0.093
	61	0.000	0.018	0.915	0.033	0.034
	121	0.000	0.006	0.978	0.007	0.009
	181	0.000	0.002	0.993	0.003	0.002
	241	0.000	0.000	1.000	0.000	0.000

Table 6: Selection estimators from $\mathcal{S} = \{T_{1/3}, T_{2/3}\}$ vs. $T_{1/3}$ and $T_{2/3}$.

n	10 p.o.t. better	equal	all p.o.t. better
19	0.406	0.485	0.109
31	0.755	0.094	0.151
61	0.923	0.001	0.076
121	0.967	0.000	0.033
181	0.990	0.000	0.010
241	0.992	0.000	0.008

Table 7: Selection estimators from $\mathcal{S} = \{T_{1/3}, T_{2/3}\}$: 10 points of time vs. all points of time.

$\mathcal{S}$	$n = 19$	$n = 31$	$n = 61$	$n = 121$
$\{T_{1/3}, T_{2/3}\}$	0.471	0.682	0.854	0.963
$\{T_{5/12}, T_{7/12}\}$	0.540	0.776	0.924	0.988
$\{T_{1/3}, T_{1/2}, T_{2/3}\}$	0.457	0.709	0.875	0.981
$\{T_{5/12}, T_{1/2}, T_{7/12}\}$	0.531	0.814	0.926	0.993
$\{T_{1/3}, T_{5/12}, T_{7/12}, T_{2/3}\}$	0.472	0.719	0.888	0.979
$\{T_{1/3}, T_{5/12}, T_{1/2}, T_{7/12}, T_{2/3}\}$	0.471	0.738	0.886	0.984

Table 8: Selection from up to five estimators: Percentage (calculated from 1 000 simulation runs) of outperforming the arithmetic mean.

n	$\mathcal{S}$	At least as good as ... estimators 5	4	3	2	1	0
19	$\{T_{1/3}, T_{2/3}\}$	—	—	—	0.148	0.726	0.126
	$\{T_{5/12}, T_{7/12}\}$	—	—	—	0.148	0.726	0.126
	$\{\overline{X}, \overline{Y}\}$	—	—	—	0.148	0.726	0.126
	$\{T_{1/3}, T_{1/2}, T_{2/3}\}$	—	—	0.068	0.414	0.442	0.076
	$\{T_{5/12}, T_{1/2}, T_{7/12}\}$	—	—	0.089	0.451	0.345	0.115
	$\{T_{1/3}, T_{5/12}, T_{7/12}, T_{2/3}\}$	—	0.041	0.186	0.500	0.207	0.066
	$\{T_{1/3}, T_{5/12}, T_{1/2}, T_{7/12}, T_{2/3}\}$	0.036	0.159	0.289	0.260	0.184	0.072
31	$\{T_{1/3}, T_{2/3}\}$	—	—	—	0.161	0.785	0.054
	$\{T_{5/12}, T_{7/12}\}$	—	—	—	0.161	0.785	0.054
	$\{\overline{X}, \overline{Y}\}$	—	—	—	0.161	0.785	0.054
	$\{T_{1/3}, T_{1/2}, T_{2/3}\}$	—	—	0.139	0.594	0.243	0.024
	$\{T_{5/12}, T_{1/2}, T_{7/12}\}$	—	—	0.144	0.674	0.139	0.043
	$\{T_{1/3}, T_{5/12}, T_{7/12}, T_{2/3}\}$	—	0.126	0.350	0.431	0.077	0.016
	$\{T_{1/3}, T_{5/12}, T_{1/2}, T_{7/12}, T_{2/3}\}$	0.119	0.352	0.286	0.167	0.063	0.013
61	$\{T_{1/3}, T_{2/3}\}$	—	—	—	0.661	0.325	0.014
	$\{T_{5/12}, T_{7/12}\}$	—	—	—	0.661	0.325	0.014
	$\{\overline{X}, \overline{Y}\}$	—	—	—	0.661	0.325	0.014
	$\{T_{1/3}, T_{1/2}, T_{2/3}\}$	—	—	0.653	0.267	0.073	0.007
	$\{T_{5/12}, T_{1/2}, T_{7/12}\}$	—	—	0.648	0.289	0.052	0.011
	$\{T_{1/3}, T_{5/12}, T_{7/12}, T_{2/3}\}$	—	0.651	0.165	0.156	0.025	0.003
	$\{T_{1/3}, T_{5/12}, T_{1/2}, T_{7/12}, T_{2/3}\}$	0.650	0.137	0.132	0.058	0.019	0.004
121	$\{T_{1/3}, T_{2/3}\}$	—	—	—	0.892	0.108	0.000
	$\{T_{5/12}, T_{7/12}\}$	—	—	—	0.892	0.108	0.000
	$\{\overline{X}, \overline{Y}\}$	—	—	—	0.892	0.108	0.000
	$\{T_{1/3}, T_{1/2}, T_{2/3}\}$	—	—	0.913	0.076	0.011	0.000
	$\{T_{5/12}, T_{1/2}, T_{7/12}\}$	—	—	0.898	0.097	0.005	0.000
	$\{T_{1/3}, T_{5/12}, T_{7/12}, T_{2/3}\}$	—	0.914	0.050	0.033	0.003	0.000
	$\{T_{1/3}, T_{5/12}, T_{1/2}, T_{7/12}, T_{2/3}\}$	0.919	0.043	0.031	0.006	0.001	0.000

Table 9: Selection from up to five estimators: Percentage (calculated from 1 000 simulation runs) of outperforming the estimators in $\mathcal{S}$.