Brandenburg
University of Technology
Cottbus - Senftenberg

Working Paper

2020/02

About the theory of simple average forecast combinations

Mathias Käso (Brandenburg University of Technology, Germany), mathias.kaeso@b-tu.de

Chair of Energy Economics
Brandenburg University of Technology Cottbus-Senftenberg
Prof. Dr. Felix Müsgens
http://www.b-tu.de/fg-energiewirtschaft/

TABLE OF CONTENTS

ABSTRACT

In this article, we theoretically derive and analyze several important properties of the simple average combination. Starting from an insightful expression for the mean squared error, we formulate the necessary and sufficient conditions for the occurrence of a perfect forecast. Additionally, we derive further performance characteristics of the simple average such as lower and upper limits for the mean squared error. The mean squared error of the combination is practically always between these limits with a certain probability to exceed the best individual forecast. Subsequently, we introduce and analyze two criteria, which determine when the combined forecast is superior to the individual forecasts. Finally, we test and confirm the most important results in several Monte Carlo experiments as well as with data from the M4 Forecast Competition. These partly new findings contribute to a complete theoretical understanding of the simple average combination and are useful to improve the prediction quality.

Keywords: *analytical investigation, equal weight combination, general performance, mean squared error*

JEL classification: *C53*

1 INTRODUCTION

It is well known that forecast combinations are a proven procedure to reach a higher forecast quality for a given set of competing forecasts. Since the initial work of Bates and Granger (1969), many alternative methods have been developed to combine a given set of forecasts. The spectrum includes e.g. simple linear combinations like the simple average or combination weights expressed by the relative mean squared error (Stock and Watson, 2001) as well as more sophisticated approaches based on regression weights (Granger and Ramanathan, 1984), time-varying regime-switching weights (Guidolin and Timmermann, 2009) or Bayesian model averaging (Madigan and Raftery, 1994; Jackson and Karlsson, 2004). A basic overview about forecast combinations can be found e.g. in Timmermann (2006) or Moral-Benito (2015) and the references therein. The resulting combined forecasts are usually better than the best individual forecast. This follows e.g. from extensive studies which analyses the performance of forecast combinations in different areas (Clemen, 1989), for thousands of time series (Makridakis and Hibon, 2000) as well as for linear and non-linear models (Stock and Watson, 2001; Stock and Watson, 2004; Marcellino, 2004).

Despite the significant effort in developing new sophisticated methods, it has been found that the simple average (weights equal to $1/N$, where N is the number of forecasts) is often hard to beat by advanced combination procedures. The simple average is frequently used and even recommended in literature because of its good performance, its simple structure, and the fact that no estimation of model parameters is necessary (see e.g. Granger and Jeon, 2004; Genre et al., 2013). Therefore, it is important to understand better the theoretical principles that determine its outstanding performance.

We can divide the studies on the performance of the simple average into two main streams. On one hand, we have the analysis of the good and partly superior performance of the simple average compared to other, especially more sophisticated combination approaches. This is formally known as the "forecast combination puzzle" (Stock and Watson, 2004). Here the majority of literature (c.f. for example Clemen and Winkler, 1986; Hendry and Clements, 2004 or more recently Smith and Wallis, 2009; Cleaskens et al., 2016; Magnus and De Luca, 2016, and Chan and Pauwels, 2018) investigate the effects of estimation issues and errors occurring in the estimation of model parameters, usually needed in more advanced procedures. They concluded that estimated parameters are rarely optimal and often inferior to fixed weights. In contrast, Timmermann (2006) and Elliott (2011) follow a slightly different argumentation. They consider the loss caused by the estimation process in more advanced combination methods and compare it to the possible gain when using optimal models. Elliott (2011) found that the optimal parameters are often similar to the weights of the simple average, caused by certain, frequently occurring properties of the individual forecasts. It follows that the gain compared to fixed equal weights is small and can be outweighed by the loss from estimation errors.

On the other hand, we are interested in the general performance of the simple average, especially in comparison to the individual forecasts from which the combination is generated. Unfortunately, from the theoretical point of view, little is known about the mathematical properties of the simple average combination of forecasts. That concerns issues such as:

- Is there a perfect forecast[1] possible with simple average combination and when yes, under which conditions?
- How bad and how good can the simple average forecast be and when is the simple average reaching these limits?
- When and why can the simple average beat each of the individual forecasts from which it is calculated?
- How can we explain the empirical findings based on theoretical results?

Recently, Chan and Pauwels (2018) have investigated especially the last two points theoretically by introducing a useful variant of the common matrix formalism. Nevertheless, why the simple average often outperforms the best individual forecast is still not really understood. They unfortunately deliver only a satisfactory answer for the unrealistic assumption of completely uncorrelated forecast errors, which can barely explain the empirical findings.

In this article, we want to provide more insights into the questions above and further properties of the simple average combination. For this purpose, we derive a formalism, which is effective and sufficient to discuss analytically several important properties of the simple average combination of $N \in \mathbb{N}$ different unbiased individual forecasts in terms of the mean squared error (MSE). Based on these partly new insights, the article contributes to complete the theory of simple average combinations. The results are useful for further investigations and can help to improve the forecast accuracy through an exact analytical understanding.

The remainder of the article is organized as follows. In Section 2, we provide the analytical discussion, divided in four parts beginning with the derivation of an insightful expression for

[1] In the sense of mean squared error.

the MSE of the simple average combination (Section 2.1). Subsequently, we investigate the possibility of perfect forecasts (Section 2.2) as well as the limits on the performance (Section 2.3) in simple average combinations. In Section 2.4, we introduce and theoretically analyze the criteria, which have to be fulfilled so that the simple average is superior to the best individual forecast. In Section 3, we apply these findings to explain the superior performance of the simple average, frequently observed in the empirical literature. We illustrate our results by a Monte Carlo simulation as well as by an application to the forecasts of the M4 Forecast Competition in Section 4 and 5. Section 6 conclude the article.

2 THE THEORETICAL ANALYSIS OF THE AVERAGE COMBINATION

In the following sections, we provide the theoretical framework for the analyzation of simple average combinations. We derive a formalism which is useful and sufficient to discuss analytically several properties of the simple average combination under quadratic loss. We deliberately avoid the common matrix formalism existing in most articles, because the compact and abstract representation sometimes hinders the recognition of deeper connections. The changed perspective leads to several, partly new insights that helps to understand the general performance of simple average combinations.

2.1 The MSE of the average combination

Let $\varphi_t, t = 1, \dots, \tau$ be the observed time series of interest and let $\varphi_{it}, i = 1, \dots, N$ be N forecasts for the observation φ_t. Then we formulate the simple average combination φ_N at time t by

$$\varphi_{Nt} = \frac{1}{N}\sum_{i=1}^{N} \varphi_{it}. \tag{1}$$

When we define the forecast error of forecast i at time t as $\Delta\varphi_{it} := \varphi_{it} - \varphi_t$, then with (1) we can derive the error of the combined forecast:

$$\begin{aligned}\Delta\varphi_{Nt} &= \left(\frac{1}{N}\sum_{i=1}^{N}\varphi_{it}\right) - \varphi_t = \left(\frac{1}{N}\sum_{i=1}^{N}\Delta\varphi_{it} + \frac{1}{N}\sum_{i=1}^{N}\varphi_t\right) - \varphi_t \\ &= \frac{1}{N}\sum_{i=1}^{N}\Delta\varphi_{it} + \left(\sum_{i=1}^{N}\frac{1}{N} - 1\right)\varphi_t = \frac{1}{N}\sum_{i=1}^{N}\Delta\varphi_{it}.\end{aligned} \tag{2}$$

From (2) it follows, that the error of the simple average combination is given by the average of the individual forecast errors. This result is typical for each combination approach in which the weights sum up to one. Let now $\Delta\boldsymbol{\varphi}_N \coloneqq (\Delta\varphi_{N1}, \dots, \Delta\varphi_{N\tau})^T$ and $\Delta\boldsymbol{\varphi}_i \coloneqq (\Delta\varphi_{i1}, \dots, \Delta\varphi_{i\tau})^T$, then the mean value $\mathrm{E}(\Delta\boldsymbol{\varphi}_N)$ can be written as

$$E(\Delta\boldsymbol{\varphi}_N) = \frac{1}{N}\sum_{i=1}^{N} E(\Delta\boldsymbol{\varphi}_i). \tag{3}$$

If we use the standard assumption $\mathrm{E}(\Delta\boldsymbol{\varphi}_i) = 0$, $\forall i$ (unbiased forecasts), then it results from (3) that $E(\Delta\boldsymbol{\varphi}_N) = 0$ as well. In this case the $MSE(\varphi_N) = \mathrm{E}(\Delta\boldsymbol{\varphi}_N)^2 + V(\Delta\boldsymbol{\varphi}_N)$ of the combined forecast is identical with the variance $V(\Delta\boldsymbol{\varphi}_N)$ of the combined errors. Therefore, we find for the MSE:

$$MSE(\varphi_N) = \sigma_N^2 = \frac{1}{N^2}\left[\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{j=2}^{N}\sum_{i=1}^{j-1}\sigma_{ij}\right]. \tag{4}$$

Here, we have introduced the common abbreviations $\sigma_N^2 \coloneqq V(\Delta\boldsymbol{\varphi}_N)$ for the variance of the combined errors, $\sigma_i^2 \coloneqq V(\Delta\boldsymbol{\varphi}_i)$ for the variance of the individual errors as well as $\sigma_{ij} \coloneqq CV(\Delta\boldsymbol{\varphi}_i, \Delta\boldsymbol{\varphi}_j)$ for their covariance. From (4), we can finally obtain an insightful representation of the variance of the combined forecast:

$$\sigma_N^2 = \frac{1}{N}\bar{\sigma}_V + \frac{N-1}{N}\bar{\sigma}_{CV}. \tag{5}$$

In (5) $\bar{\sigma}_V$ is the average of all N error variances σ_i^2 of the individual forecasts, given by

$$\bar{\sigma}_V \coloneqq \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2. \tag{6}$$

Analogous $\bar{\sigma}_{CV}$ is the average of all $N(N-1)/2$ covariances σ_{ij} between the errors of the individual forecasts, which we can write as

$$\bar{\sigma}_{CV} \coloneqq \frac{2}{N(N-1)}\sum_{j=2}^{N}\sum_{i=1}^{j-1}\sigma_{ij}. \tag{7}$$

For finite variances σ_i^2 and finite covariances σ_{ij}, the average variance $\bar{\sigma}_V$ and the average covariance $\bar{\sigma}_{CV}$ are also finite, with values $\bar{\sigma}_V \in [\sigma_{min}^2, \sigma_{max}^2]$ and $\bar{\sigma}_{CV} \in [\sigma_{min}, \sigma_{max}]$. Here are $\sigma_{min}^2 = \min\{\sigma_i^2 | \, i \in [1, N]\}$ and $\sigma_{max}^2 = \max\{\sigma_i^2 | \, i \in [1, N]\}$ the minimum and maximum error variance as well as $\sigma_{min} = \min\{\sigma_{ij} | j \in [2, N] \wedge i \in [1, j-1]\}$ and $\sigma_{max} = \max\{\sigma_{ij} | \, j \in [2, N] \wedge i \in [1, j-1]\}$ the minimum and maximum covariance. This is even true, when we consider an infinite number N of forecasts. In this case (6) and (7) are infinite but convergent series, with limit values again in $[\sigma_{min}^2, \sigma_{max}^2]$ and $[\sigma_{min}, \sigma_{max}]$ (the proof is based on the direct comparison test). We can see from (5), that the error variance of the combined forecast is the weighted average of the average variance $\bar{\sigma}_V$ and the average covariance $\bar{\sigma}_{CV}$, where the weights sum up to one:

$$\frac{1}{N} + \frac{N-1}{N} = 1. \tag{8}$$

Equation (5) reduces the error variance from a function of a large number of statistical parameters to an expression with only two relevant statistical parameters. This is useful for the following discussions.

2.2 The perfect forecast in average combinations

We start our discussion of perfect forecasts in simple average combinations with an academic case and for a finite number N of forecasts. If we assume the unrealistic but simple and often discussed scenario of pairwise completely uncorrelated errors ($\rho_{ij} = 0$, $\forall ij$ with ρ_{ij} as the correlation between errors i and j), then, according to (5), the covariance part vanishes and the variance of the combined forecast is given by

$$\sigma_N^2 = \frac{1}{N}\bar{\sigma}_V. \tag{9}$$

Obviously, under the strong assumption of completely uncorrelated errors and a finite $\bar{\sigma}_V$ (see Section 2.1), σ_N^2 is decreasing with increasing the number N of forecasts combined, leading to a perfect forecast for $N \to \infty$. The same result was previously found by Chan & Pauwels (2018) discussing the conditions for the superiority of the simple average compared to the best individual forecast. They derived the result using a matrix formalism framework. Here it follows simply from (5). Note there is an additional case resulting in (9), not discussed by Chan & Pauwels (2018). The result as well as the argumentation is also valid in case of a perfect cancelling out of the positive and negative error covariances of the individual forecasts, because then too $\bar{\sigma}_{CV} = 0$ [see (7)].

However, in order to explain the empirical findings, it is necessary to analyze a more realistic scenario. Therefore, we have to discuss the entire problem given in (5). From this equation, we find for $\bar{\sigma}_{CV}$:

$$\bar{\sigma}_{CV} \geq -\frac{1}{N-1}\bar{\sigma}_V. \tag{10}$$

The relation follows directly from the condition that the variance σ_N^2 of the combined error must be greater than or equal to zero. Furthermore, we see that a simple average combination can deliver a perfect forecast only in case of equality in (10). That means the average covariance $\bar{\sigma}_{CV}$ has to take a certain non-positive value between $-\bar{\sigma}_V$ for $N = 2$ and zero for $N \to \infty$, depending on the respective number of forecasts and their average variance. This is the necessary and sufficient condition for a perfect forecast assuming that the average covariance is not vanishing. In practice, this condition is unlikely to be fulfilled, because it is well known that in most cases the errors of the individual forecast are positively correlated, leading to a positive average covariance $\bar{\sigma}_{CV}$. The reason for this general observation is that frequently similar model approaches as well as similar data sets are used for the forecasting process. This results in similar error structures which, of course, show positive correlations (for a deeper discussion see e.g. Elliott, 2011).

Now, we want to consider (5) for an increasing N. Since the weights in (5) sum up to one, it follows that σ_N^2 is more and more dominated by the covariances. For example let $N = 10$, then only 10 % of $\bar{\sigma}_V$ but 90 % of $\bar{\sigma}_{CV}$ contributes to σ_N^2. For the limit $N \to \infty$, we even have

$$\sigma_N^2 = \bar{\sigma}_{CV}. \tag{11}$$

That means that σ_N^2 is completely independent from a direct contribution of the error variances and purely determined by the covariances of the forecast errors. In addition, from

(10) as well as (11), we find the unexpected result that for $N \to \infty$ the average covariance cannot be negative ($\bar{\sigma}_{CV} \geq 0$). It is again necessary to ensure a non-negative variance σ_N^2 in (11). However, the most important aspect is that even under the extreme assumption $N \to \infty$ the error variance of the combined forecast is normally not vanishing (c.f. also the last paragraph of Section 2.1). This result is not in accordance with Chan & Pauwels (2018) and will lead to significantly different conclusions about the performance of the simple average combination, especially in comparison to the best individual forecast (see Sections 2.3 and 2.4).

From this first analysis, it follows that it is beneficial to use a large number of forecasts that are as diverse as possible in simple average combinations. The large number of forecasts ensures a small contribution of the individual error variances to the combined error variance [see (5)]. The diversification means that we use forecasts, which have a higher probability for small or even negative error covariances. Such condition is at least reducing the dominant part $\bar{\sigma}_{CV}$ in (5) or even give a negative sign to $\bar{\sigma}_{CV}$ which both lead to a smaller σ_N^2. Therefore, we have here a theoretical explanation for the "conventional belief" (c.f. Chan & Pauwels, 2018) that diversified forecasts are more beneficial in forecast combinations.

However, these conditions are nearly non-existent in reality. Rather, we have a finite, often a small number of forecast as well as highly correlated errors with not rarely similar error variances (see e.g. Elliott, 2011; Elliott & Timmermann 2016 or the discussion in Granger and Jeon, 2004). Based on these circumstances, we can neither expect something near to a perfect forecast from simple average combinations nor use this argument to explain the frequent outperformance of the simple average combination compared to the individual forecasts, as Chan and Pauwels (2018) proposed it. Therefore, we have to find a more satisfactory answer, which is the aim of the following sections.

2.3 Limits of performance in average combinations

To analyze the superiority of the simple average compared to the best individual and therefore to all individual forecasts, first we analyze the general bounds of the average combination performance. Using the limit $\sigma_i^2 \to \sigma_{max}^2, \forall i$, we obtain from (5)[2]

$$\sigma_N^2 = \frac{1}{N}\bar{\sigma}_V + \frac{N-1}{N}\bar{\sigma}_{CV} \leq \left[\frac{1}{N} + \frac{N-1}{N}\bar{\rho}\right]\sigma_{max}^2 := \sigma_{N,max}^2, \tag{12}$$

when

$$\bar{\rho} \geq \bar{\rho}_{min,0} = \frac{1}{N-1}\left[\frac{\bar{\sigma}_V}{\sigma_{max}^2} - 1\right] + \frac{\bar{\sigma}_{CV}}{\sigma_{max}^2}. \tag{13}$$

Here $\bar{\rho}$ is the average of all $N(N-1)/2$ correlations ρ_{ij} between the N forecast errors $\Delta\boldsymbol{\varphi}_i$, analogously defined to (7). From (10), we can determine the smallest possible $\bar{\rho}$. If $\sigma_i^2 \to \sigma_{max}^2, \forall i$, then the right hand side of (10) is minimal with $-\sigma_{max}^2/(N-1)$. Simultaneously, the left hand side simplifies to $\sigma_{max}^2\bar{\rho}$ resulting in

[2] More precise, we consider the limit of a function: $\lim_{\sigma_i^2 \to \sigma_{max}^2, \forall i} [\sigma_N^2(\sigma_1^2, \dots, \sigma_N^2)] = \sigma_N^2(\sigma_{max}^2, \dots, \sigma_{max}^2)$.

$$\bar{\rho} \geq \bar{\rho}_{min} = -\frac{1}{N-1}. \tag{14}$$

If we combine the minimum number of forecasts ($N = 2$), then $\bar{\rho}_{min} = -1$. For the opposite case, $N \to \infty$, we find that $\bar{\rho}_{min} \to 0$. Therefore, we have together $\bar{\rho}_{min} \in [-1,0]$. However, from (14), it follows, that already for a small number N of forecasts, we can have only a slightly negative average correlation. Since the correlation ρ_{ij} between the errors is mostly positive, we can even expect a positive $\bar{\rho}$ in the majority of cases.

If $\rho_{ij} \geq 0$, $\forall ij \Rightarrow \bar{\rho} \geq 0$ then the inequality (12) is generally true and $\sigma^2_{N,max}$ is an upper bound. The upper bound $\sigma^2_{N,max}$ is maximized for $\bar{\rho} = 1$ which follows, when all errors are perfectly correlated ($\rho_{ij} = 1$, $\forall ij$). Since both weights sum up to one [see (8)], we find then

$$\sigma^2_N \leq \sigma^2_{max}. \tag{15}$$

That means the error variance of the combined forecast cannot be higher than the highest error variance of the individual forecasts.

In general, we cannot assume that always $\rho_{ij} \geq 0$, even if this is likely for the majority of cases. In this case, the inequality (12) holds only if $\bar{\rho} \geq \bar{\rho}_{min,0}$ with a certain value $\bar{\rho}_{min,0} \in [\bar{\rho}_{min}, 1]$ according to (13). Here, we are interested in the largest $\bar{\rho}_{min,0}$ defined as $\bar{\rho}_{min,l}$ which the average correlation $\bar{\rho}$ has to match or even exceed. For a corresponding analysis of (13), it is helpful to divide it into two steps. In the first step, we assume once again for the correlations $\rho_{ij} \geq 0$, $\forall ij$, which is a common property for forecast errors and therefore the more relevant case. Here we should find a value of $\bar{\rho}_{min,l}$ for which $\bar{\rho} \geq \bar{\rho}_{min,0}$ is always satisfied, because we already expect the validity of (12) for this assumption. With the limit $\sigma^2_i \to \sigma^2_{max}$, $\forall i$, we can maximize $\bar{\rho}_{min,0}$ and simultaneously rewrite the average covariance as $\bar{\sigma} = \sigma^2_{max}\,\bar{\rho}$. Using these results in (13) we have for the largest possible $\bar{\rho}_{min,0}$,

$$\bar{\rho}_{min,l} = \bar{\rho}. \tag{16}$$

Under the assumption that $\rho_{ij} \geq 0$, $\forall ij \Rightarrow \bar{\rho} \geq 0$ and according to (13) and (16) as well as the related discussion, we can summarize the following result:

$$\bar{\rho} \geq \bar{\rho}_{min,l} \geq \bar{\rho}_{min,0}. \tag{17}$$

Indeed, since $\bar{\rho}_{min,0}$ has the maximum value $\bar{\rho}_{min,l} = \bar{\rho}$, the average correlation always fulfils the condition $\bar{\rho} \geq \bar{\rho}_{min,0}$. Therefore $\sigma^2_{N,max}$ in (12) is always an upper bound for the error variance of the combined forecast, if $\rho_{ij} \geq 0$, $\forall ij$.

In the second step, we additionally allow for negative correlations ρ_{ij} between the errors. Then, the limit $\sigma^2_i \to \sigma^2_{max}$, $\forall i$ maximizes again the first term of the sum in (13). The second term of the sum however, is not necessarily increasing to his maximum value. We can see this, when we rewrite this term as a weighted sum of the correlations. With (7) it follows

$$\frac{\bar{\sigma}_{CV}}{\sigma_{max}^2} = \frac{2}{N(N-1)} \sum_{j=2}^{N} \sum_{i=1}^{j-1} \frac{\sigma_i \sigma_j}{\sigma_{max}^2} \rho_{ij}. \tag{18}$$

When we assume now, that the possible negative correlations in (18) have mostly small weights $\sigma_i \sigma_j / \sigma_{max}^2 \ll 1$, while the positive correlations have weights close to one. Then, after applying the limiting process $\sigma_i^2 \to \sigma_{max}^2$, $\forall i$, the negative correlations receive more weight ($\sigma_i \sigma_j / \sigma_{max}^2 \to 1$), whereas the weights of the positive correlations are only slightly changed. In such a case, it is possible that $\bar{\sigma}_{CV} / \sigma_{max}^2$ has a smaller value after the limiting process, because the negative correlations have now an increased contribution to the sum. This leads to a reduced summation value. Therefore, it is not guaranteed that the limit $\sigma_i^2 \to \sigma_{max}^2$, $\forall i$ in (13) delivers the largest $\bar{\rho}_{min,0}$.

For a deeper analyses of this point, we now introduce $\Delta\sigma_i^2 \coloneqq \sigma_{max}^2 - \sigma_i^2 \geq 0$ and $\Delta\sigma_{ij} \coloneqq \sigma_{ij,max} - \sigma_{ij}$ with $\sigma_{ij,max} \coloneqq \sigma_{max}^2 \rho_{ij}$ ($\Delta\sigma_{ij} \geq 0$ if $\rho_{ij} \geq 0$ and $\Delta\sigma_{ij} < 0$ if $\rho_{ij} < 0$) as the changes of the variance and the covariance terms caused by the limiting process. Then from (6), we can write for the limit $\sigma_i^2 \to \sigma_{max}^2$, $\forall i$:

$$\lim_{\sigma_i^2 \to \sigma_{max}^2, \forall i} [\bar{\sigma}_V] = \frac{1}{N} \sum_{i=1}^{N} \sigma_{max}^2 = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 + \Delta\sigma_i^2 = \bar{\sigma}^2 + \frac{1}{N} \sum_{i=1}^{N} \Delta\sigma_i^2. \tag{19}$$

Without loss of generality, we can count the $N(N-1)/2$ individual correlations in (7) by a new index $(i,j) \to k$ with $k = 1, \dots, K$ and $K \coloneqq N(N-1)/2$. Then the limiting process $\sigma_i^2 \to \sigma_{max}^2$, $\forall i$ delivers for (7)

$$\lim_{\sigma_i^2 \to \sigma_{max}^2, \forall i} [\bar{\sigma}_{CV}] = \frac{1}{K} \sum_{k=1}^{K} \sigma_{max}^2 \rho_k = \frac{1}{K} \sum_{k=1}^{K} \sigma_k + \Delta\sigma_k = \bar{\sigma} + \frac{1}{K} \sum_{k=1}^{K} \Delta\sigma_k. \tag{20}$$

Based on these two expressions, it is possible to investigate when $\bar{\rho}_{min,l}$, which results from $\bar{\rho}_{min,0}$ and the limit $\sigma_i^2 \to \sigma_{max}^2$, $\forall i$, is smaller than $\bar{\rho}_{min,0}$. Using (19) and (20), we can rewrite the inequality $\bar{\rho}_{min,l} \leq \bar{\rho}_{min,0}$:

$$\begin{aligned}
\lim_{\sigma_i^2 \to \sigma_{max}^2, \forall i} [\bar{\rho}_{min,0}] &= \frac{1}{N-1} \left[\frac{\lim_{\sigma_i^2 \to \sigma_{max}^2, \forall i} [\bar{\sigma}_V]}{\sigma_{max}^2} - 1 \right] + \frac{\lim_{\sigma_i^2 \to \sigma_{max}^2, \forall i} [\bar{\sigma}_{CV}]}{\sigma_{max}^2} \\
&= \frac{1}{N-1} \left[\frac{\bar{\sigma}_V + \frac{1}{N} \sum_{i=1}^{N} \Delta\sigma_i^2}{\sigma_{max}^2} - 1 \right] + \frac{\bar{\sigma}_{CV} + \frac{1}{K} \sum_{k=1}^{K} \Delta\sigma_k}{\sigma_{max}^2} \\
&= \bar{\rho}_{min,0} + \frac{1}{N(N-1)} \sum_{i=1}^{N} \frac{\Delta\sigma_i^2}{\sigma_{max}^2} + \frac{1}{K} \sum_{k=1}^{K} \frac{\Delta\sigma_k}{\sigma_{max}^2} \\
&= \bar{\rho}_{min,l} \leq \bar{\rho}_{min,0}.
\end{aligned} \tag{21}$$

With $K \coloneqq N(N-1)/2$ and from the last two lines in (21) it follows that the sum over all changes has to be equal to or smaller than zero, i.e.

$$\sum_{i=1}^{N} \Delta\sigma_i^2 + 2\sum_{k=1}^{K} \Delta\sigma_k \leq 0. \tag{22}$$

To determine the influence of the negative correlations and thus the influence of the negative covariances, we divide the sum over k into a negative and positive part. To this end, we assign the L terms with negative correlations as $\Delta\sigma_k^-$and the $K - L$ terms with positive correlations as $\Delta\sigma_k^+$.[3] Now, we obtain $\bar{\rho}_{min,l} \leq \bar{\rho}_{min,0}$ when

$$\sum_{i=1}^{N} \Delta\sigma_i^2 + 2\sum_{k=1}^{K-L} \Delta\sigma_k^+ \leq 2\sum_{k=1}^{L} |\Delta\sigma_k^-|. \tag{23}$$

The inequality (23) means that after the limiting process the summation value over all changes $\Delta\sigma_i^2 \geq 0$ of the variances and all changes $2\Delta\sigma_k^+ \geq 0$ of the positive covariances has to be smaller or equal to the summation value over all changes $2|\Delta\sigma_k^-| \geq 0$ of the negative covariances.

In practice, there is a low probability that (23) can be fulfilled by a realistic set of forecasts and therefore $\bar{\rho}_{min,l}$ according to (16) is still the maximum of $\bar{\rho}_{min,0}$. The reason being, is that the occurrence of negative error correlations is rare, resulting in $L \ll N + (K - L)$. Meaning we can only expect a small number of positive terms on the right-hand side of (23) compared to the left-hand side. In addition, these terms are related to error correlations, which are typically small. But then the terms $|\Delta\sigma_k^-|$ are small as well, because $\left|\Delta\sigma_{ij}\right| \coloneqq \left|\sigma_{ij,max} - \sigma_{ij}\right| = |(\sigma_{max}^2 - \sigma_i\sigma_j)\rho_{ij}|$ is small, if $\left|\rho_{ij}\right| \ll 1$. That is especially valid, since there is no reason that the coefficient $(\sigma_{max}^2 - \sigma_i\sigma_j) \geq 0$ in $\Delta\sigma_k^-$ are systematically large, whereas these coefficients are systematically small in $\Delta\sigma_k^+$ [see the argumentation following (18)]. Together, it is likely that we have a predominance of the terms on the left-hand side in (23) and it is $\bar{\rho}_{min,l} \geq \bar{\rho}_{min,0}$. Therefore, we assume that the condition (17) is fulfilled in nearly all practical cases, even when not all error correlations are strictly positive (see also Section 4 and 5). Then $\sigma_{N,max}^2$ from (12) is an upper bond, which restricts the values of σ_N^2 upwards. This restriction of σ_N^2 can be further specified as we show in the Appendix. There we find, that σ_N^2 is in addition always smaller than or equal to the average of all variances of the individual forecasts defined in (6). Therefore, it is more precise

$$\begin{aligned} \sigma_N^2 &\leq \sigma_{N,max}^2 \leq \bar{\sigma}_V, && \text{if } \bar{\rho} \leq \bar{\rho}_x, \\ \sigma_N^2 &\leq \bar{\sigma}_V \leq \sigma_{N,max}^2, && \text{if } \bar{\rho} > \bar{\rho}_x. \end{aligned} \tag{24}$$

Here $\bar{\rho}_x$ is given by (43) and results from the condition $\bar{\sigma}_V = \sigma_{N,max}^2(\bar{\rho})$, which represents the intersection between the constant $\bar{\sigma}_V$ and the function $\sigma_{N,max}^2(\bar{\rho})$ (see also Figure 1).

Analogously to (12), we can find a lower bound for the combined forecast. From (5), we obtain

$$\sigma_N^2 = \frac{1}{N}\bar{\sigma}_V + \frac{N-1}{N}\bar{\sigma}_{CV} \geq \left[\frac{1}{N} + \frac{N-1}{N}\bar{\rho}\right]\sigma_{min}^2 \coloneqq \sigma_{N,min}^2. \tag{25}$$

[3] In both cases, we introduce a further new numeration, so that $k = 1, \ldots, L$ respectively $k = 1, \ldots, K - L$.

Here, the inequality (25) holds true in general when

$$\bar{\rho} \leq \bar{\rho}_{max,0} = \frac{1}{N-1}\left[\frac{\bar{\sigma}_V}{\sigma_{min}^2} - 1\right] + \frac{\bar{\sigma}_{CV}}{\sigma_{min}^2}. \tag{26}$$

Analogues to the analysis of (13), we want to divide the analysis of $\bar{\rho}_{max,0}$ in two parts. Firstly, we consider again that all correlations are positive, $\rho_{ij} \geq 0$, $\forall ij$. Then the smallest $\bar{\rho}_{max,0}$ possible, defined as $\bar{\rho}_{max,s}$, can be found for the limit $\sigma_i^2 \to \sigma_{min}^2$, $\forall i$, leading to $\bar{\rho}_{max,s} = \bar{\rho}$ [see also (16)]. In that case, it is always

$$\bar{\rho} \leq \bar{\rho}_{max,s} \leq \bar{\rho}_{max,0} \tag{27}$$

and $\sigma_{N,min}^2$ is the lower bound for the combined forecast σ_N^2 according to (25). In this context, we can directly discuss an interesting special case of positive correlations. Let $\rho_{ij} = 1$, $\forall ij \ \Rightarrow \ \bar{\rho} = 1$. Then we obtain from (25)

$$\sigma_N^2 \geq \sigma_{min}^2, \qquad \text{if} \ \ \rho_{ij} = 1, \forall ij. \tag{28}$$

This means, that a simple average combination of forecasts with pairwise perfectly positive correlated errors cannot be better than the forecast with the smallest error variance σ_{min}^2. Even if we expect highly positive correlated forecast errors, this case should be very rare. Therefore, in general an outperformance of the best individual forecast should be at least possible.

Now, we expand again the analysis on the possibility of negative correlations. Then, similar to the discussion of (13), we cannot ensure, that the limit $\sigma_i^2 \to \sigma_{min}^2$, $\forall i$ minimizes $\bar{\rho}_{max,0}$ and therefore that $\bar{\rho} \leq \bar{\rho}_{max,0}$. While the first term of the sum in (26) is unchanged by allowing for negative correlations, we can again focus the discussion to the second term of the sum. The term can be written similarly to (18) as a weighted sum of the correlations:

$$\frac{\bar{\sigma}_{CV}}{\sigma_{min}^2} = \frac{2}{N(N-1)} \sum_{j=2}^{N} \sum_{i=1}^{j-1} \frac{\sigma_i \sigma_j}{\sigma_{min}^2} \rho_{ij}. \tag{29}$$

In the above case, all weights $\sigma_i \sigma_j / \sigma_{min}^2$ are larger than one with $\sigma_i \sigma_j / \sigma_{min}^2 \to 1$ for $\sigma_i^2 \to \sigma_{min}^2, \forall i$. Independent of this distinguishing property to (18), we can still construct theoretically cases in which $\bar{\sigma}_{CV}/\sigma_{min}^2$ increases after applying the limiting process. But then, there are values of $\bar{\rho}_{max,0}$ with $\bar{\rho}_{max,0} \leq \bar{\rho}_{max,s}$ and $\bar{\rho}_{max,s}$ is not the smallest $\bar{\rho}_{max,0}$ anymore.

To analyze the point, we can proceed similarly to the investigation of (17). To this end, we introduce the variance and covariance changes as $\Delta\sigma_i^2 \coloneqq \sigma_{min}^2 - \sigma_i^2 \leq 0$ and $\Delta\sigma_{ij} \coloneqq \sigma_{ij,min} - \sigma_{ij}$ with $\sigma_{ij,min} \coloneqq \sigma_{min}^2 \rho_{ij}$ ($\Delta\sigma_{ij} < 0$ if $\rho_{ij} > 0$ and $\Delta\sigma_{ij} \geq 0$ if $\rho_{ij} \leq 0$).[4] With calculations analogous to (19) - (23), we can show that

[4] We use here the same nomenclature for the changes $\Delta\sigma_i^2$ and $\Delta\sigma_{ij}$ to keep it simple and because we do not expect any confusion.

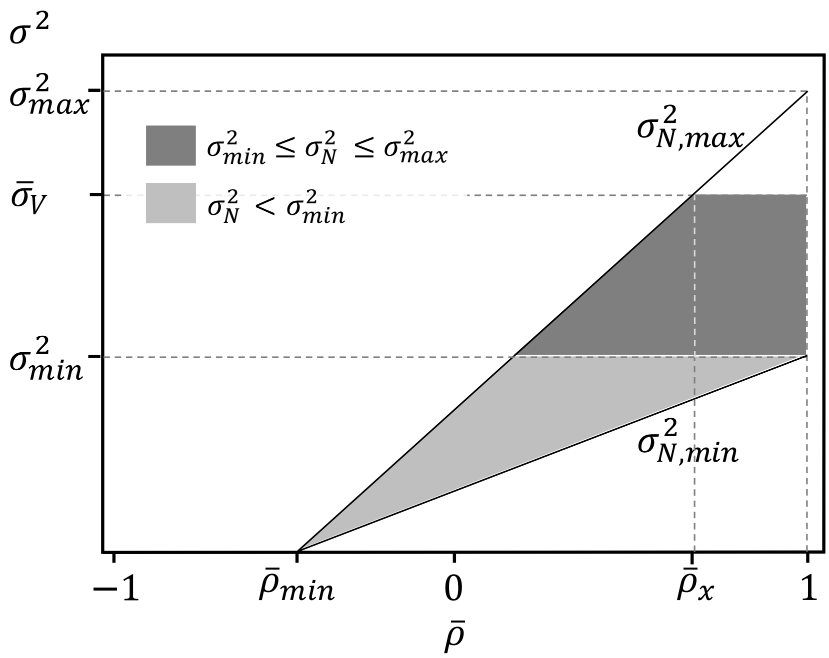

Figure 1: Schematic representation of the allowed area of the error variance σ_N^2 (medium and dark grey areas) as a function of $\bar{\rho}$. Also depicted are $\sigma^2_{N,min}$ and $\sigma^2_{N,max}$ (black solid lines) as well as $\bar{\rho}_{min}$, $\bar{\rho}_x$, σ^2_{min}, $\bar{\sigma}_V$, and σ^2_{max} (dashed lines).

$$\sum_{i=1}^{N} |\Delta\sigma_i^2| + 2\sum_{k=1}^{K-L} |\Delta\sigma_k^+| \leq 2\sum_{k=1}^{L} \Delta\sigma_k^-. \tag{30}$$

Based on this condition and following the same reasoning as in (23), it is likely that the inequality (27) can also be fulfilled for negative correlations in the majority of practical cases (see also in Section 4 and 5). From here on, we can therefore assume that the conditions (17) and (27) are valid and $\sigma^2_{N,max}$ as well as $\sigma^2_{N,min}$ are the upper and lower bounds of the error variance of the forecast combination with $\sigma^2_{N,min} \leq \sigma^2_N \leq \sigma^2_{N,max}$.

Figure 1 summarizes the results of the previous investigation. There, we have schematically represented the area in which we can expect the variance σ_N^2 of the combined forecast (medium and dark grey areas) as a function of the average correlation $\bar{\rho}$. The possible σ_N^2 are confined by the black lines $\sigma^2_{N,max}$ and $\sigma^2_{N,min}$ as well as $\bar{\sigma}_V$ according to (12), (24) and (25) where $\sigma^2_{N,max}$ and $\sigma^2_{N,min}$ can be rewritten as linear functions of $\bar{\rho}$. In the dark grey area σ_N^2 of the combined forecast cannot beat the best individual forecast with variance σ^2_{min}, but also never exceeds the variance σ^2_{max} of the worst individual forecast [see (15)]. Especially for $\bar{\rho} = 1$ there is no combined forecast that can beat the best individual forecast [see (28)]. In the medium grey area, the combined forecast is better than each of the individual forecasts and reaches a perfect prediction for $\bar{\rho} = \bar{\rho}_{min}$ according to (14). This can be seen e.g., when we use (14) in (12). Then we find that $\sigma_N^2 \leq 0 \Rightarrow \sigma_N^2 = 0$, because σ_N^2 is non-negative. In addition, we observe an increasing share of the medium grey area compared to the dark grey area for $\bar{\rho} \to \bar{\rho}_{min}$. Therefore, we can note, that we find an increasing probability for the superiority of the combined forecast with a decreasing average correlation. That is again an argument to use diverse individual forecasts. A deeper analysis of the conditions under which we can expect that the $\sigma_N^2 < \sigma_{min}^2$ will be addressed in the next section.

As the last point in this section, we want to quantify the error variance limits in which we can expect the error variance of the combined forecast as well as the maximum improvement potential compared to best individual forecast.

The lower line $\sigma_{N,min}^2$ and the upper line $\sigma_{N,max}^2$ in Figure 1 are the theoretical minimum and maximum for the variance σ_N^2 of the combined forecast. The minimum $\sigma_{N,min}^2$ can only be reached by σ_N^2, if all error variances σ_i^2 are equal [see (25)]. Then, it is $\sigma_i^2 = \sigma_{min}^2 = \sigma_{max}^2$, $\forall i$ and the grey areas in Figure 1 are contracting to a line. In this case, it follows that σ_N^2 cannot be higher than the σ_i^2 of the individual forecasts. In all other cases σ_N^2 lies between $\sigma_{N,min}^2$ and $\sigma_{N,max}^2$. In Figure 2 we have depicted the ratio $R \coloneqq \sigma_{N,min}^2/\sigma_{min}^2 = \sigma_{N,max}^2/\sigma_{max}^2$ based on (12) and (25) as a function of $\bar{\rho}$. The function values in this figure can be multiplied with σ_{min}^2 or σ_{max}^2 to find the theoretical minimum or maximum variance of σ_N^2 for each set of given forecasts. For example, in the case of three forecasts with an average error correlation $\bar{\rho} = 0$, the variance σ_N^2 cannot be lower than $(1/3)\sigma_{min}^2$ and not be higher than $(1/3)\sigma_{max}^2$.

This case is highlighted in Figure 2. Moreover, we see that the dependence of R on the number of forecasts N is decreasing with increasing N. For $N > N_0$ and N_0 sufficiently large, we can express approximately the theoretical minimum variance $\sigma_{N,min}^2$ of the combined forecast simply by:

$$\sigma_{N,min}^2 \cong \bar{\rho}\sigma_{min}^2 . \tag{31}$$

From this approximation, we see the minimum possible variance $\sigma_{N,min}^2$ of the combined forecast is determined by a fraction of σ_{min}^2, where the fraction is simply given by $\bar{\rho} \in [\bar{\rho}_{min}, 1]$. Note, that the approximation can only be used, when $\bar{\rho} \geq 0$, since $\sigma_{N,min}^2$ is non-negative. But this condition is almost always fulfilled for large N [see (14)]. Analogously, we have $\sigma_{N,max}^2 \cong \bar{\rho}\sigma_{max}^2$ if $\bar{\rho} \leq \bar{\rho}_x$ for the theoretical maximum of σ_N^2. For $\bar{\rho} > \bar{\rho}_x$ the upper bound is given by $\bar{\sigma}_V$.

The maximum improvement potential compared to best individual forecast can be expressed from the vertical distance between $\sigma_{min}^2 - \sigma_{N,min}^2 \coloneqq \Delta\sigma_{min}^2$ in relation to σ_{min}^2 (see Figure 1):

$$\frac{\Delta\sigma_{max}^2}{\sigma_{min}^2} = \frac{N-1}{N}[1-\bar{\rho}] \cong 1-\bar{\rho}, \qquad \text{if} \quad N > N_0, \tag{32}$$

with $\bar{\rho} \geq \bar{\rho}_{min}$. The more forecasts we consider and the less the associated forecast errors are correlated on average, the greater the improvement can be through the combination. If we have only a few forecasts and the associated forecast errors are highly correlated, then it is not excluded theoretically to be better than the best individual forecast, but the expected gain from the combination is rather small. Finally, the question arises under which conditions and probability we can access this potential for improvement what is addressed in the next section.

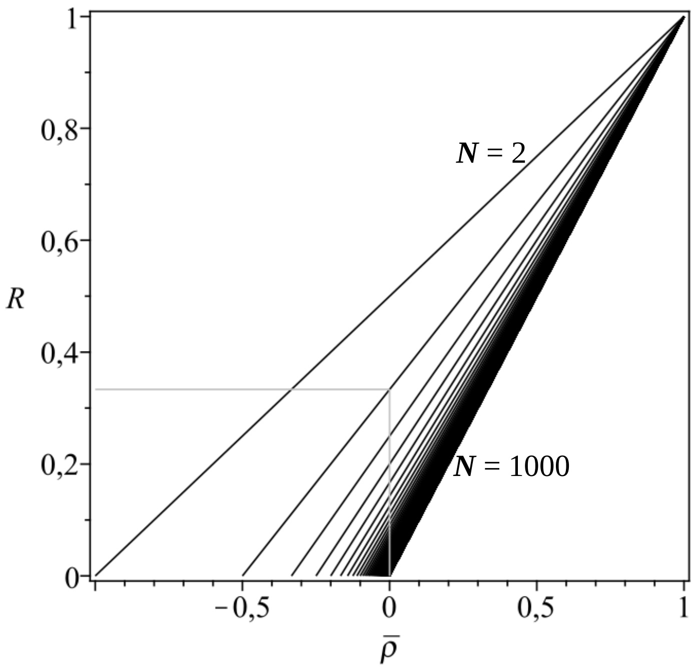

Figure 2: Representation of the ratio R as function of the average correlation $\bar{\rho}$ for different numbers of forecasts N. The grey lines indicate the case of three forecasts with an average correlation of zero (see also the description in the text).

2.4 Superiority of the simple average to the individual forecasts

In this section, we want to derive suitable criteria for a superiority of the combined forecast over the best individual forecast. To this end, firstly we consider again the unrealistic case of pairwise completely uncorrelated errors ($\rho_{ij} = 0, \forall ij$). Then, according to (9), the covariance part vanishes and the error variance of the combined forecast is given by $\sigma_N^2 = (1/N)\bar{\sigma}_V$. Under this strong assumption the necessary and sufficient condition for the superiority of a certain individual forecast i compared to the simple average can be formulated as

$$\sigma_i^2 < \sigma_N^2 = \frac{1}{N}\bar{\sigma}_V. \tag{33}$$

If and only if the error variance σ_i^2 of the individual forecast i is smaller than the average error variance $\bar{\sigma}_V$ over all forecasts divided by the number N of forecasts combined, the individual forecast will outperform the average combination. It follows that σ_i^2 has to be at least half of the average error variance, when we consider the minimum number of forecasts ($N = 2$) included in a combination. As discussed before, in case of completely uncorrelated errors and a finite $\bar{\sigma}_V$ (see Section 2.1), σ_N^2 is decreasing with increasing N, which leads to a perfect forecast for $N \to \infty$. That means, it is highly unlikely for an individual forecast to outperform the simple average combination, if N is large and all forecast errors are uncorrelated. Chan & Pauwels (2018) already discussed this result, but it follows here simply from (5). Note here too, there is an additional case, not considered by Chan & Pauwels (2018). Since the condition as well as the argumentation always holds true when $\bar{\sigma}_{CV} = 0$, it is also valid in case of a perfect cancelling out of positive and negative error covariances of the individual forecasts.

In the more realistic case of arbitrary, often positively correlated forecast errors, Chan & Pauwels (2018) propose the same explanation given above for the superiority of the simple average combination. However, as was shown in Section 2.2, we do not find in this case that σ_N^2 is vanishing for $N \to \infty$. Therefore, we cannot expect to beat each of the individual forecasts by simply increasing the number of forecasts combined. Moreover, such argument would not explain the superiority of the simple average in the regular case of a finite, not that big number of individual forecast. Therefore, we need to find another theoretical explanation for the often better performance compared to the individual forecasts. To this end, we investigate when the upper bound $\sigma_{N,max}^2$ is less than or equal to the best individual forecast with variance σ_{min}^2. Then, according to (12), we set

$$\left[\frac{1}{N} + \frac{N-1}{N}\bar{\rho}\right] \sigma_{max}^2 \leq \sigma_{min}^2 . \tag{34}$$

Here $\bar{\rho}$ is restricted to a limiting value $\bar{\rho}_{lim}$, so that $\bar{\rho} \leq \bar{\rho}_{lim} \leq 1$, which ensures that the left hand side of (34) is less than or equal to σ_{min}^2. The value of $\bar{\rho}_{lim}$ can be calculated from the equality case in (34).[5] It is

$$\bar{\rho}_{lim} = \frac{1}{N-1}\left[N\sigma_{mm}^2 - 1\right]. \tag{35}$$

Here, we have introduced the ratio $\sigma_{mm}^2 = \sigma_{min}^2 / \sigma_{max}^2$ with $\sigma_{mm}^2 \in (0,1]$, where we explicitly exclude $\sigma_{min}^2 = 0$. In this case, we would already have a perfect forecast and no forecast combination is needed anymore. The upper limit $\bar{\rho}_{lim}$ for the average correlation is only depending on the number of forecasts as well as the minimum and maximum error variances, which are confining the variances of all forecasts included in the combination. When N is additionally large, $\bar{\rho}_{lim}$ is practically given by the ratio σ_{mm}^2. We have $\bar{\rho}_{lim}$ depicted in Figure 3.

For $\bar{\rho} \leq \bar{\rho}_{lim}$ and under the assumptions of Section 2.3, we find σ_N^2 in the triangular light grey area in which $\sigma_N^2 \leq \sigma_{min}^2$ always holds true. However, we can also see that $\bar{\rho} \leq \bar{\rho}_{lim}$ is a relative strict criterion, because it is excluding a relative high amount of possible values for σ_N^2 coming from the medium grey area where we have $\sigma_N^2 \leq \sigma_{min}^2$ as well. Therefore, we want to additionally introduce a second criterion based on the average variance $\sigma_{avg}^2 := (\sigma_{min}^2 + \sigma_{max}^2)/2$ which is more related to typical values of σ_N^2. Substituting σ_{max}^2 in (34) by σ_{avg}^2, we get a new limit value for the average correlation $\bar{\rho}$, and we expect that $\sigma_N^2 \leq \sigma_{min}^2$ in most of the cases. It follows that

$$\bar{\rho}_{avg} = \frac{1}{N-1}\left[N\sigma_{ma}^2 - 1\right] = \frac{1}{N-1}\left[2N\frac{\sigma_{mm}^2}{1+\sigma_{mm}^2} - 1\right]. \tag{36}$$

Here, we have defined the ratio $\sigma_{ma}^2 = \sigma_{min}^2/\sigma_{avg}^2$. Since $\sigma_{ma}^2 \geq \sigma_{mm}^2$, we always get $\bar{\rho}_{avg} \geq \bar{\rho}_{lim}$. For the right hand side of (36), we have used $\sigma_{max}^2 = \sigma_{min}^2 / \sigma_{mm}^2$ in σ_{ma}^2 to express $\bar{\rho}_{avg}$ in terms of the ratio σ_{mm}^2. In this way, it is easy to compare $\bar{\rho}_{lim}$ with $\bar{\rho}_{avg}$.

[5] This value is given by the intersection between σ_{min}^2 and $\sigma_{N,min}^2$ in Figure 3.

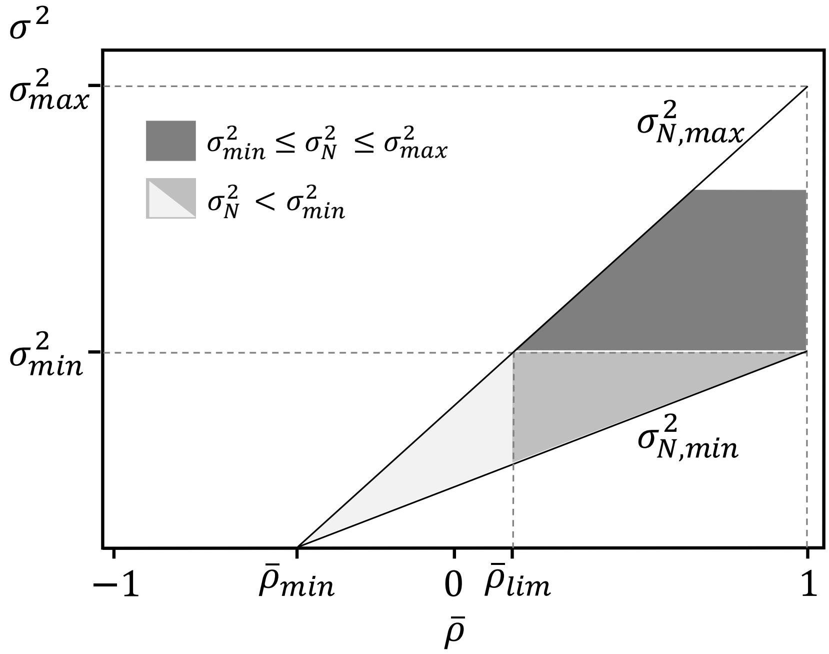

Figure 3: Schematic representation of the allowed area of the error variance σ_N^2 (light, medium and dark grey areas) as a function of $\bar{\rho}$. Also depicted are $\sigma_{N,min}^2$ and $\sigma_{N,max}^2$ (black solid lines) as well as $\bar{\rho}_{min}$, $\bar{\rho}_{lim}$, σ_{min}^2 and σ_{max}^2 (dashed lines).

In Figure 4, we have represented the relative position of $\bar{\rho}_{avg}$ and $\bar{\rho}_{lim}$. We see that $\bar{\rho}_{avg}$ is clearly less restrictive, because it is now allowing to include more of the typical variance values around the average line with $\sigma_N^2 \le \sigma_{min}^2$. These values are indicated by the dashed triangle. However, for the less restrictive $\bar{\rho}_{avg}$, it is possible that we include in rare cases some combined forecasts with $\sigma_N^2 > \sigma_{min}^2$ (see the upper corner of the dashed triangle). Nevertheless, in most cases $\bar{\rho} \le \bar{\rho}_{avg}$ should be a sufficient criterion to decide if we can expect superiority of the combined forecast over the individual forecasts (see also Section 4 and 5). In case of $\bar{\rho} > \bar{\rho}_{avg}$ it is more and more unlikely to find a superior combined forecast with $\sigma_N^2 \le \sigma_{min}^2$ although it is not excluded.

In Figure 5 we have represented the possible $\bar{\rho}_{lim}$ and $\bar{\rho}_{avg}$ given in (35) and (36) as functions of the ratio $\sigma_{mm}^2 = \sigma_{min}^2 / \sigma_{max}^2$ for different numbers N of forecasts. We see, that in case of using the minimum number of forecasts in the combination ($N = 2$), all values between zero and one are possible for $\bar{\rho}_{lim}$ and $\bar{\rho}_{avg}$. For increasing N, this range is in both cases shrinking to only positive $\bar{\rho}_{lim}$ and $\bar{\rho}_{avg}$, in accordance to the fact, that no average correlation can be smaller than $\bar{\rho}_{min}$ [see (14)]. As discussed before, we see that $\bar{\rho}_{avg} > \bar{\rho}_{lim}$ with the exception $\sigma_{mm}^2 = 0$ and $\sigma_{mm}^2 = 1$, where we have $\bar{\rho}_{avg} = \bar{\rho}_{lim}$.

From the condition (17) we know, that $\sigma_N^2 \le \sigma_{N,max}^2$, which practically holds true for each $\bar{\rho}$ determined by the set of forecasts considered. Then we can finally summarize [together with (35)] that

$$\bar{\rho} \le \bar{\rho}_{lim} \Rightarrow \sigma_N^2 \le \sigma_{N,max}^2 \le \sigma_{min}^2. \tag{37}$$

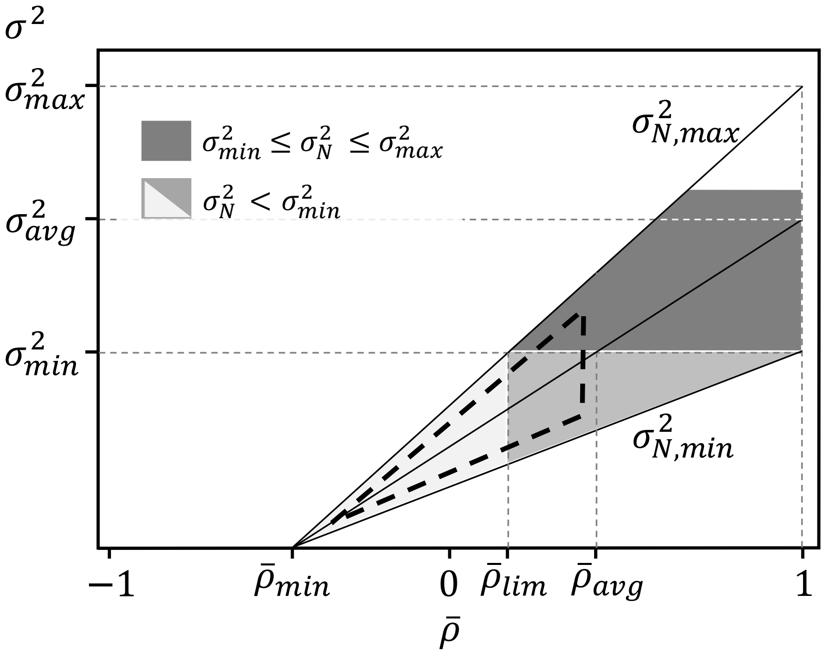

Figure 4: Schematic representation of the allowed area of the error variance σ_N^2 (light, medium and dark grey areas) as a function of $\bar{\rho}$. Also depicted are $\sigma^2_{N,min}$ and $\sigma^2_{N,max}$ (black solid lines) as well as $\bar{\rho}_{min}$, $\bar{\rho}_{lim}$, $\bar{\rho}_{avg}$, σ^2_{min} and σ^2_{max} (dashed lines). In addition, typical values for σ_N^2 are indicated as a dashed triangle.

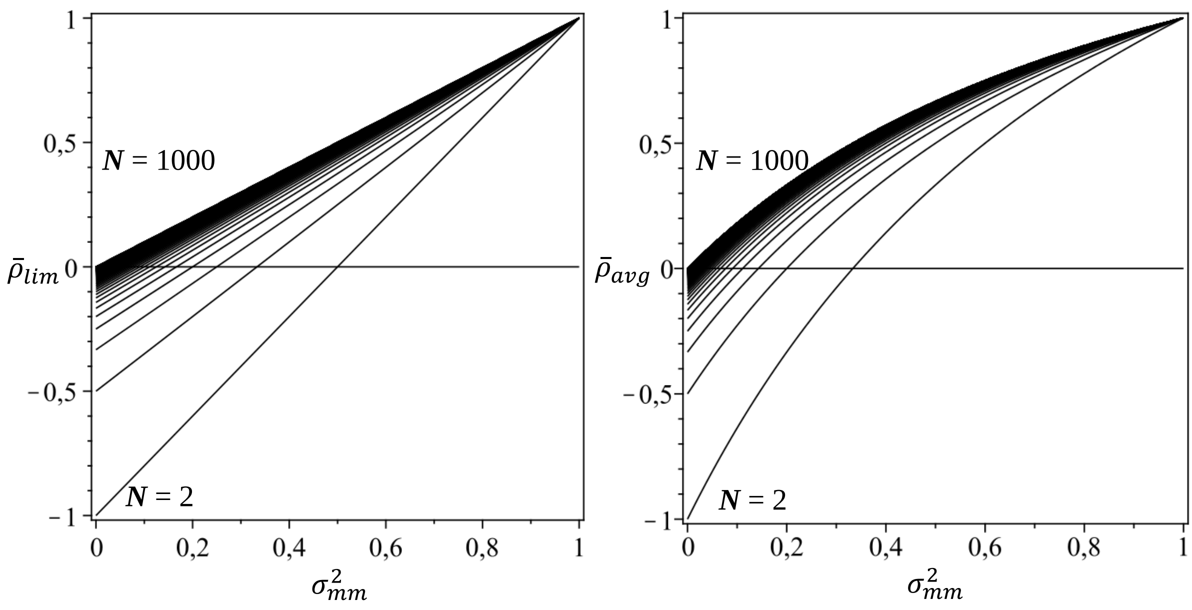

Figure 5: Representation of the $\bar{\rho}_{lim}$ and $\bar{\rho}_{avg}$ as functions of the ratio $\sigma^2_{mm} = \sigma^2_{min}/\sigma^2_{max}$ for different numbers N of forecasts.

While (37) is always true under our weak assumptions, we expect in most cases for the error variance of combined forecasts $\sigma_N^2 \leq \sigma^2_{min}$ when $\bar{\rho} \leq \bar{\rho}_{avg}$ (see Figure 4 as well as Section 4 and 5).

3 EXPLANATION OF THE EMPIRICAL FINDINGS

We want to use the theoretical results of the previous sections to explain the empirical finding, that simple average combinations often outperform the best individual forecast. To this end, we have depicted in Figure 6 the behavior of σ_N^2 as a function of $\bar{\rho}$ for an increased ratio $\sigma_{mm}^2 \in (0,1]$ which occurs when $\sigma_{min}^2 \rightarrow \sigma_{max}^2$ (indicated in the figure by the black arrow). Compared to Figure 4 and according to Figure 5, we observe in this case an increasing $\bar{\rho}_{lim}$ and $\bar{\rho}_{avg}$ leading to a higher probability that $\bar{\rho}$ is satisfying the condition (37) as well as the less restrictive condition $\bar{\rho} \leq \bar{\rho}_{avg}$. From this observation, we can note: The closer σ_{mm}^2 is to one, the greater the probability that the combined forecast will outperform the best individual forecast.

According to Elliott (2011) the error variances of the individual forecasts are indeed often very similar, resulting in σ_{mm}^2 frequently close to one (see also the discussion in Granger and Jeon, 2004). This leads to a high positive value of $\bar{\rho}_{lim}$ and $\bar{\rho}_{avg}$. Then, as previously showed, it is very likely that the average correlation of the forecast errors $\bar{\rho}$ is smaller than $\bar{\rho}_{lim}$, fulfilling the condition (37). In addition there is an even higher probability for $\bar{\rho} \leq \bar{\rho}_{avg}$. In this case, we find directly, that the combined forecast is better than the best individual forecast. But even, when σ_{mm}^2 deviates stronger from one, we can expect a superiority of the simple average forecast when the average correlation of the errors is not too large. Here we find again, that it is beneficial to use diverse individual forecasts in simple average combinations. Both arguments together are our explanation for the superiority of the average combination compared to the individual forecasts often found in literature. This effect is mostly based on the widespread properties of the individual forecasts and does not follow from a better performance of the combination with increasing the number N of forecasts included [see (11)]. On the contrary, we observe a weak dependency of the criteria $\bar{\rho} \leq \bar{\rho}_{lim}$ and $\bar{\rho} \leq \bar{\rho}_{avg}$ from N, when N is not too small (see Figure 5). That means from $N \cong 10$ a higher number of forecasts has only a minor contribution to the superiority of the simple average. More important is a high ratio $\sigma_{mm}^2 \in (0,1]$ of the minimum and maximum error variance of the individual forecasts.

Conversely, that means too, that in case of $\sigma_{mm}^2 \ll 1$ ($\sigma_{min}^2 \ll \sigma_{max}^2$), we have $\bar{\rho}_{lim} \cong \bar{\rho}_{avg} \rightarrow \bar{\rho}_{min}$ [see (14)]. Then it is unlikely that $\bar{\rho}$ is small enough, since $\bar{\rho}$ is normally determined by higher positive correlations ρ_{ij}. Here, at least the best individual forecast can outperform the average combination.

4 NUMERICAL ILLUSTRATION

In this section, we illustrate our theoretical findings by a simple numerical experiment. This experiment is based on sets of six correlated, normally distributed error samples. Both assumptions are common and often fairly well fulfilled by realistic error samples. To generate such kind of forecast errors in our simulation, we start with a set of six normally distributed errors, $\boldsymbol{\varepsilon}' = (\varepsilon_1', \dots, \varepsilon_6')^T$ with mean vector $\boldsymbol{\mu}' = (\mu_1', \dots, \mu_6')^T = \mathbf{0}$ ($\mathbf{0}$ is the (6×1) zero vector) and covariance matrix $\boldsymbol{\Sigma}' = \boldsymbol{I}$ ($\boldsymbol{I}$ is the (6×6) identity matrix). Then, we can generate

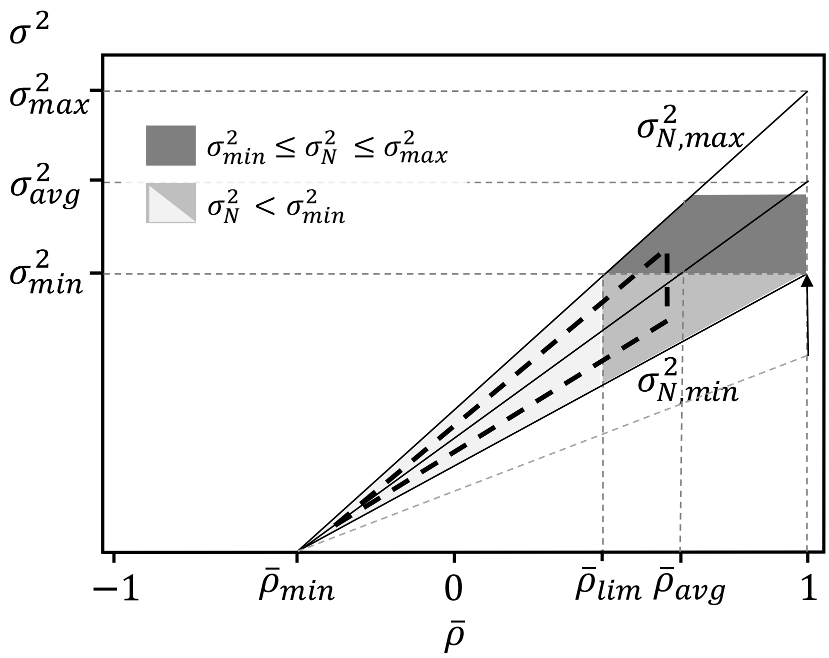

Figure 6: Schematic representation of the allowed area of the error variance σ_N^2 (light, medium and dark grey areas) for the shift $\sigma^2_{min} \rightarrow \sigma^2_{max}$ as a function of $\bar{\rho}$. Also depicted are $\sigma^2_{N,min}$ and $\sigma^2_{N,max}$ (black solid lines) as well as $\bar{\rho}_{min}$, $\bar{\rho}_{lim}$, $\bar{\rho}_{avg}$, σ^2_{min} and σ^2_{max} (dashed lines). In addition, typical values for σ_N^2 are indicated as a dashed triangle.

a set of correlated errors $\boldsymbol{\varepsilon}$ with certain $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by using the transformation (see also Diebold & Mariano 1995)[6]

$$\boldsymbol{\varepsilon} = \boldsymbol{\mu} + \boldsymbol{C}\boldsymbol{\varepsilon}'. \tag{38}$$

Here $\boldsymbol{C}$ is a lower triangular matrix resulting from a Cholesky decomposition of $\boldsymbol{\Sigma}$. In this experiment and according to the theory presented here, we choose $\boldsymbol{\mu} = \boldsymbol{0}$ [see Section 2.1]. In order to generate a large amount of different error samples $\boldsymbol{\varepsilon}$, we define two error variances in $\boldsymbol{\Sigma}$ as σ^2_{min} and σ^2_{max} and choose the remaining variances randomly between them. Furthermore, all correlations and therefore all covariances in $\boldsymbol{\Sigma}$ are also random but are restricted to a certain interval $[\rho_1, \rho_2]$. In this way, we generate different error samples $\boldsymbol{\varepsilon}$, where in general only σ^2_{min} and σ^2_{max} are the same. Finally, we combine the six errors of each sample according to Equation (2) and compute the variance of the combined error.

However, there is one problem with this kind of error generation process. Not each compilation of randomly chosen $\rho_{ij} \in [\rho_1, \rho_2]$ is allowed. For example, imagine we have two highly negative correlated error samples, then a third error sample cannot also be highly negative correlated with both of them. With one of them, the third sample should have most likely a positive correlation. In such scenarios the Cholesky decomposition of $\boldsymbol{\Sigma}$ is nonexistent. For more than three error samples, the discussion of allowed correlations shows a strongly increasing complexity. In order to not further extend the theoretical part of this

[6] In our numerical example, we do not generate errors with serial correlation, because we believe that serial correlation is mainly a problem for combination methods in which the model coefficients has to be estimated (see Gauss–Markov theorem).

article, we pragmatically use only samples in which the decomposition and therefore the transformation according to (38) exists.

In Figure 7, the results of this experiment are presented. The Subfigures A – D in Figure 7 represent the error variance σ_N^2 of the combined forecast as a function of $\bar{\rho}$ for different ratios σ_{mm}^2 (A: $\sigma_{mm}^2 = 4/5$, B: $\sigma_{mm}^2 = 3/5$, C: $\sigma_{mm}^2 = 2/5$, D: $\sigma_{mm}^2 = 1/5$). The individual correlations ρ_{ij} are randomly chosen from $[-0.2, 0.7]$ to explicitly allow for negative correlations. The average correlation $\bar{\rho}$ is then calculated from these 15 individual correlations ρ_{ij} of each error set generated.

As expected from the theoretical analyses in cases A – D, all σ_N^2 are confined between $\sigma_{N,min}^2$ and $\sigma_{N,max}^2$ (grey solid lines), even though we find negative individual correlations in the generated data (see discussion in Section 2.3). In A and B, all $\bar{\rho}$ are smaller than the strict limit $\bar{\rho}_{lim}$ and indeed it is always the case that $\sigma_N^2 \leq \sigma_{min}^2$. In Subfigure C only roughly half of the σ_N^2 have an average correlation $\bar{\rho}$ smaller than $\bar{\rho}_{lim}$. Still, most σ_N^2 are smaller than σ_{min}^2. Here, we have an example that in some cases $\bar{\rho} \leq \bar{\rho}_{lim}$ is a too strict criterion for the superiority of the combined forecast. In contrast, the second criterion $\bar{\rho} \leq \bar{\rho}_{avg}$ is violated only for one single σ_N^2. Therefore, we find that according to this criterion, the majority of σ_N^2 is smaller than σ_{min}^2. The few occurring exceptions are expected and have been discussed in Section 2.4. Finally, in Subfigure D the majority of σ_N^2 has a $\bar{\rho} > \bar{\rho}_{avg}$, and indeed we obtain now the result that most of the σ_N^2 cannot beat the variance σ_{min}^2 of the best individual forecast.

5 EMPIRICAL ILLUSTRATION

To verify empirically our most important theoretical results further, we use the data of the M4 Forecasting Competition (Makridakis et al., 2020). We investigate all 100.000 times series of the package. The package contains hourly, daily, weekly, monthly, quarterly and yearly time series and related forecasts for different forecasting problems and forecasting horizons. A detailed description of the data can be found in Makridakis et al. (2020). For each time series are 25 forecasts available. In our investigation, we combine the 5, 10 und 15 best forecasts of each time series according to (1) and compute σ_N^2, σ_{min}^2, σ_{max}^2, $\bar{\sigma}_V$, $\sigma_{N,min}^2$ and $\sigma_{N,max}^2$ as well as $\bar{\rho}$, $\bar{\rho}_{lim}$ and $\bar{\rho}_{avg}$. Subsequently, we analyze these results with respect to our theoretical findings. The results are represented in Table 1 and Table 2 where we have calculated the percentage of forecasts fulfilling our theoretical restrictions to σ_N^2 in dependence on the properties of $\bar{\rho}$ for the three different combinations as well as various types of times series.

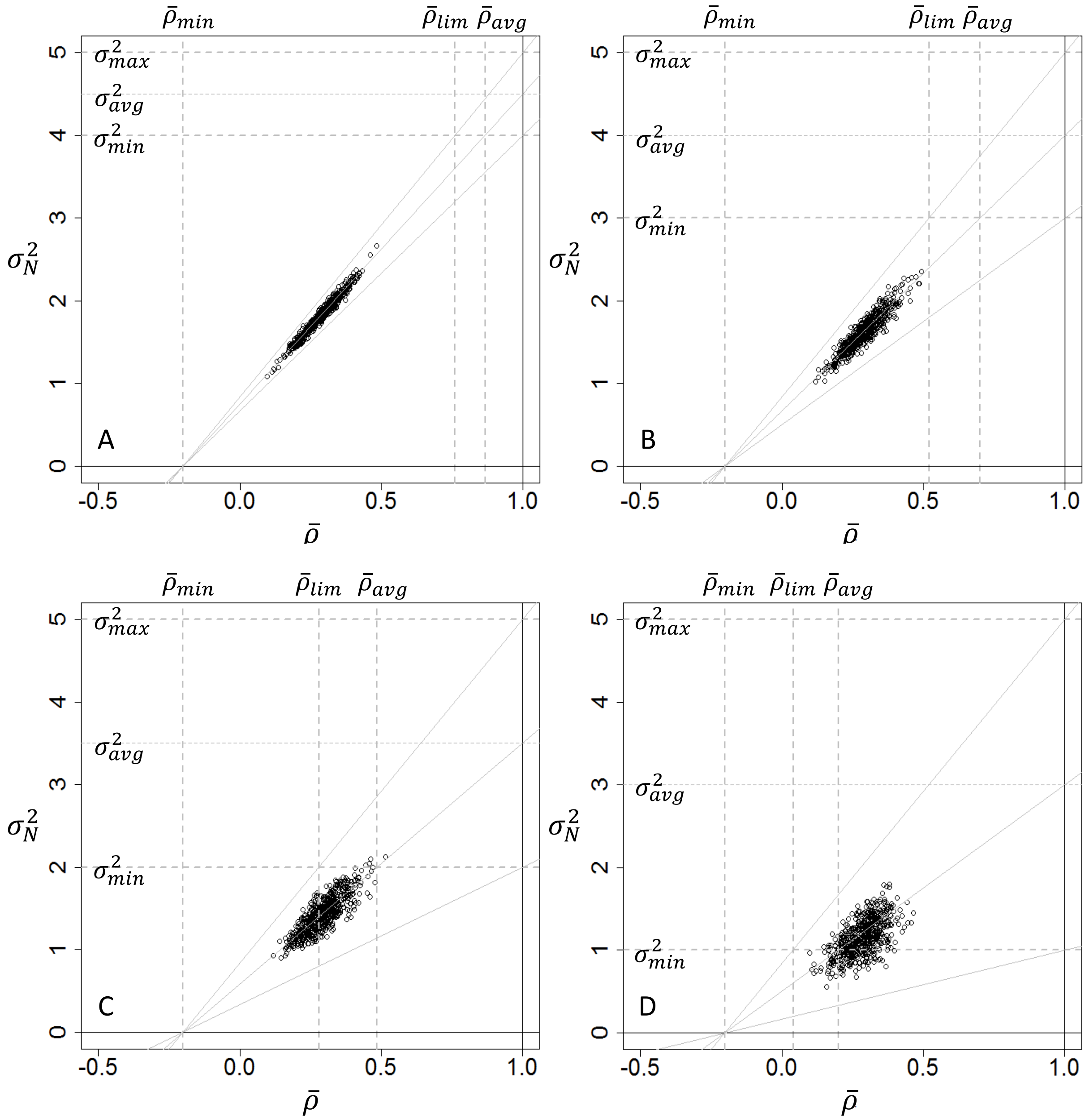

Figure 7: Representation of the error variance σ_N^2 based on a variety of error sets. Each set is consisting of six correlated, normally distributed error samples. The error sets are depicted for different ratios $\sigma_{mm}^2 = \sigma_{min}^2/\sigma_{max}^2$. A: $\sigma_{mm}^2 = 4/5$, B: $\sigma_{mm}^2 = 3/5$, C: $\sigma_{mm}^2 = 2/5$, D: $\sigma_{mm}^2 = 1/5$. The upper bound $\sigma_{N,max}^2$ and the lower bound $\sigma_{N,min}^2$ are represented by grey lines.

In accordance to (15) and (42), we find in all cases that $\sigma_N^2 \leq \sigma_{max}^2$ and $\sigma_N^2 \leq \bar{\sigma}_V$ which has to be true in general and is independent of $\bar{\rho}$. Moreover, we find that $\sigma_N^2 \leq \sigma_{N,max}^2$ and $\sigma_N^2 \geq \sigma_{N,min}^2$ in the majority of cases [see (12) and (25)]. Meaning that most forecast combinations follow our assumptions and $\sigma_{N,max}^2$ as well as $\sigma_{N,min}^2$ are indeed an upper and lower bound for σ_N^2 as discussed in Section 2.3. The strongest deviation from our expectation, we find for the daily time series ($\bar{\rho} > \bar{\rho}_{avg}$) where only slightly more than 90 % of the forecast combinations respecting these limits. Typically, however, the values are 100% or just below and are valid for all $\bar{\rho}$. Finally, we analyze the forecast combinations regarding $\sigma_N^2 \leq \sigma_{min}^2$. When $\bar{\rho} > \bar{\rho}_{avg}$ (see Section 2.4) it is unlikely, that the combination can outperform the best individual forecast and therefore, we expect only a small percentage of combinations with $\sigma_N^2 \leq \sigma_{min}^2$. The findings in Table 1 and Table 2 support exactly this behavior. Mostly, only a few percent or even less of the forecast combinations are superior to best individual forecast. The percentage of forecast combinations with $\sigma_N^2 \leq \sigma_{min}^2$ is strongly increasing when $\bar{\rho}$ from $\bar{\rho}_{lim} < \bar{\rho} \leq \bar{\rho}_{avg}$. We observe values between 42.86 % and 90.00 % in which higher values clearly dominate. Overall, on average 71.5 % of the combined forecasts are better than the best individual forecast if the average correlation $\bar{\rho}$ is from the interval $(\bar{\rho}_{lim}, \bar{\rho}_{avg}]$. For $\bar{\rho} \leq \bar{\rho}_{lim}$ near to all forecast combinations are superior to the best individual forecast. The few exceptions are resulting from cases in which the upper bound $\sigma_{N,max}^2$ is violated. In these cases of course, it is possible that $\sigma_N^2 > \sigma_{min}^2$ as well (see also Figure 3). However, the percentage of combinations with $\sigma_N^2 > \sigma_{min}^2$ has to be smaller than or equal to the percentage of combinations violating $\sigma_N^2 \leq \sigma_{N,max}^2$, which is obviously fulfilled in each case of our analysis.

Table 1: PERCENTAGES OF FORECAST COMBINATIONS FULFILLING THE RESTRICTIONS FOR σ_N^2 IN DEPENDANCE ON $\bar{\rho}$ FOR THE 5, 10, AND 15 BEST FORECASTS COMBINED.

		Forecasts combined: 5			Forecasts combined: 10			Forecasts combined: 15		
		$\bar{\rho} \leq \bar{\rho}_{lim}$	$\bar{\rho} > \bar{\rho}_{lim}$ $\bar{\rho} \leq \bar{\rho}_{avg}$	$\bar{\rho} > \bar{\rho}_{avg}$	$\bar{\rho} \leq \bar{\rho}_{lim}$	$\bar{\rho} > \bar{\rho}_{lim}$ $\bar{\rho} \leq \bar{\rho}_{avg}$	$\bar{\rho} > \bar{\rho}_{avg}$	$\bar{\rho} \leq \bar{\rho}_{lim}$	$\bar{\rho} > \bar{\rho}_{lim}$ $\bar{\rho} \leq \bar{\rho}_{avg}$	$\bar{\rho} > \bar{\rho}_{avg}$
Hourly Time Series: 414	$\sigma_N^2 \leq \sigma_{max}^2$	100.00	100.00	100.00	100.00	100.00	100.00	NA	100.00	100.00
	$\sigma_N^2 \leq \bar{\sigma}_V$	100.00	100.00	100.00	100.00	100.00	100.00	NA	100.00	100.00
	$\sigma_N^2 \leq \sigma_{N,max}^2$	100.00	100.00	100.00	100.00	100.00	100.00	NA	100.00	100.00
	$\sigma_N^2 \geq \sigma_{N,min}^2$	100.00	100.00	100.00	100.00	100.00	100.00	NA	100.00	100.00
	$\sigma_N^2 \leq \sigma_{min}^2$	100.00	81.13	9.92	100.00	79.66	11.75	NA	75.00	5.28
Daily Time Series: 4227	$\sigma_N^2 \leq \sigma_{max}^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \bar{\sigma}_V$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \sigma_{N,max}^2$	96.40	98.15	90.82	100.00	98.57	90.62	100.00	100.00	91.96
	$\sigma_N^2 \geq \sigma_{N,min}^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \sigma_{min}^2$	97.30	60.49	0.13	100.00	47.14	0.29	100.00	42.86	0.19
Weekly Time Series: 359	$\sigma_N^2 \leq \sigma_{max}^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \bar{\sigma}_V$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \sigma_{N,max}^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \geq \sigma_{N,min}^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \sigma_{min}^2$	100.00	72.73	1.85	100.00	59.26	2.75	100.00	90.00	3.16

Table 2: PERCENTAGES OF FORECAST COMBINATIONS FULFILLING THE RESTRICTIONS FOR σ_N^2 IN DEPENDANCE ON $\bar{\rho}$ FOR THE 5, 10, AND 15 BEST FORECASTS COMBINED.

		Forecasts combined: 5			Forecasts combined: 10			Forecasts combined: 15		
		$\bar{\rho} \leq \bar{\rho}_{lim}$	$\bar{\rho} > \bar{\rho}_{lim}$ $\bar{\rho} \leq \bar{\rho}_{avg}$	$\bar{\rho} > \bar{\rho}_{avg}$	$\bar{\rho} \leq \bar{\rho}_{lim}$	$\bar{\rho} > \bar{\rho}_{lim}$ $\bar{\rho} \leq \bar{\rho}_{avg}$	$\bar{\rho} > \bar{\rho}_{avg}$	$\bar{\rho} \leq \bar{\rho}_{lim}$	$\bar{\rho} > \bar{\rho}_{lim}$ $\bar{\rho} \leq \bar{\rho}_{avg}$	$\bar{\rho} > \bar{\rho}_{avg}$
Monthly Time Series: 48000	$\sigma_N^2 \leq \sigma_{max}^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \bar{\sigma}_V$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \sigma_{N,max}^2$	99.76	99.94	99.96	99.63	100.00	99.97	100.00	99.69	99.99
	$\sigma_N^2 \geq \sigma_{N,min}^2$	99.84	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \sigma_{min}^2$	99.84	77.11	2.19	100.00	71.52	1.33	100.00	69.29	1.01
Quarterly Time Series: 24000	$\sigma_N^2 \leq \sigma_{max}^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \bar{\sigma}_V$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \sigma_{N,max}^2$	99.61	99.80	99.97	99.07	100.00	99.98	100.00	100.00	100.00
	$\sigma_N^2 \geq \sigma_{N,min}^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \sigma_{min}^2$	99.71	76.87	2.15	100.00	71.11	1.40	100.00	75.93	1.09
Yearly Time Series: 23000	$\sigma_N^2 \leq \sigma_{max}^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \bar{\sigma}_V$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \sigma_{N,max}^2$	98.28	99.28	99.88	95.29	99.66	99.95	100.00	100.00	99.95
	$\sigma_N^2 \geq \sigma_{N,min}^2$	99.50	99.94	100.00	97.65	100.00	100.00	100.00	100.00	100.00
	$\sigma_N^2 \leq \sigma_{min}^2$	98.89	81.36	4.19	98.82	80.48	3.22	100.00	74.47	2.14

6 CONCLUSION

This article presents a theoretical analysis of several properties of simple average combinations. Our analysis is based on an insightful and effective mathematical description of the simple average combination problem, where we deliberately do not follow the common matrix formalism. This makes it possible for us to obtain deeper insights into the fundamental performance of this combination approach.

In this context, we derive the necessary and sufficient conditions for a perfect combination result, which occurs when the combined forecast has a $\sigma_N^2 = 0$. From the investigation and for the very unlikely case of perfectly uncorrelated forecast errors, we can reproduce existing results in the literature. Here the perfect forecast occurs only for an infinite number of forecasts, since then, the error variance is decreasing with increasing number of forecasts. We show additionally that the above mentioned condition and the corresponding argumentation are also valid in the case of a perfect cancelling out of positive and negative error covariances, which was not discussed before. However, such case is also very unlikely, since negative error covariances are usually rare and simultaneously close to zero.

Furthermore, in the more realistic scenario of arbitrary, often highly positive correlated errors, we provide a new condition. We show that this condition is practically impossible to be satisfied under normal circumstances. The reason again is based on the unlikely existence of negative error covariances. Therefore, we have in general only a very low probability to produce a perfect forecast from the simple average combination. Under the given assumptions here, the perfect forecast is not achievable even if we use an infinite number of individual forecasts, because in our results we do not find a vanishing error variance by simply increasing the number of forecasts. Instead, we show that the influence of the individual error variances on the error variance of the combination is indeed vanishing but the influence of the covariances, which quite fast dominate, is not. Here the dominance of the covariances occurs already for a realistic high number of forecasts.

Subsequently, we derive several further performance properties of the simple average combination. We show that the combined forecast is never inferior to the forecast with the highest error variance and is even superior to the average of all error variances of the individual forecasts. Moreover, if we combine a set of perfectly positive correlated errors, we cannot beat the best individual forecast. In all other cases, we show that the error variance of the combined forecast lies between a theoretical maximum and minimum error variance and can but does not have to be superior to all individual forecasts. However, the probability for superiority increases as the average correlation of the errors decrease. The advantages of using diverse forecasts in the simple average combination are thus confirmed.

Based on these results, we introduce two criteria for the allowed average error correlation, which will determine if a simple average combination is superior to all individual forecasts included in the combination. Both criteria, the strict and the less strict, are only dependent on the number of forecasts and the ratio between the minimum and the maximum error variance. The criteria are more admissive when the ratio is closer to one. With these criteria, we can explain the empirical finding that the simple average indeed often outperforms the best individual forecast. The effect is resulting mostly from similar error variances as a widespread property of the individual forecasts. These as well as the most important other theoretical results are empirically tested and confirmed in several Monte Carlo experiments as well as for 100.000 time series and 300.000 forecast combinations based on the data of the M4 Forecast Competition.

Finally, we conclude, that the simple average combination benefits from a large number of diverse forecasts with similar error variances. The large number of forecasts ensures a small contribution of the individual error variances to the combined error variance. The diversity of forecasts results in a higher probability for small or even negative error covariances. This leads to a reducing effect on the dominating part $\bar{\sigma}_{CV}$ in σ_N^2. Furthermore, the diversity helps to fulfil the criteria of superiority, because we can expect a smaller average correlation. Simultaneously, these criteria are more tolerant, if the error variances are similar. This allows the combined forecast to beat the best individual forecast. Conversely, if we have a set of forecasts with highly correlated errors and very different error variances the simple average is most likely not the best combination strategy. In such cases, using simply the best individual forecast or a more advanced combination approach would most probably outperform the simple average (see also the discussion of the forecast combination puzzle in Timmermann, 2006; Elliott, 2011 and Claeskens, 2016).

APPENDIX

Starting from the second term in (5) and using the generalized Triangle inequality as well as the Cauchy-Schwarz inequality, we obtain:

$$\frac{N-1}{N}\bar{\sigma}_{CV} = \frac{2}{N^2}\sum_{j=2}^{N}\sum_{i=1}^{j-1}\sigma_{ij} \leq \frac{2}{N^2}\sum_{j=2}^{N}\sum_{i=1}^{j-1}|\sigma_{ij}| \leq \frac{2}{N^2}\sum_{j=2}^{N}\sum_{i=1}^{j-1}\sqrt{\sigma_i^2\sigma_j^2}. \tag{39}$$

In the last expression on the right hand side of (39), we have the geometrical mean of σ_i^2 and σ_j^2 which is always smaller than or equal to the arithmetic mean of both quantities. Therefore, we can further write

$$\frac{N-1}{N}\bar{\sigma}_{CV} \leq \frac{2}{N^2}\sum_{j=2}^{N}\sum_{i=1}^{j-1}\frac{\sigma_i^2+\sigma_j^2}{2} = \frac{1}{N^2}\sum_{j=2}^{N}\sum_{i=1}^{j-1}\left(\sigma_i^2+\sigma_j^2\right). \tag{40}$$

Rearranging the terms of the last sums in (40) according equal indices leads to

$$\begin{aligned}\frac{N-1}{N}\bar{\sigma}_{CV} &\leq \frac{1}{N^2}\{(\sigma_1^2+\sigma_2^2)+[(\sigma_1^2+\sigma_3^2)+(\sigma_2^2+\sigma_3^2)]+\cdots\\ &\quad+[(\sigma_1^2+\sigma_N^2)+(\sigma_2^2+\sigma_N^2)+\cdots+(\sigma_{N-1}^2+\sigma_N^2)]\}\\ &= \frac{1}{N^2}\{(N-1)\sigma_1^2+(N-1)\sigma_2^2+\cdots+(N-1)\sigma_N^2\}\\ &= \frac{(N-1)}{N}\frac{1}{N}\{\sigma_1^2+\sigma_2^2+\cdots+\sigma_N^2\}\end{aligned} \tag{41}$$

$$= \frac{(N-1)}{N} \bar{\sigma}_V.$$

It follows, that $\bar{\sigma}_{CV} \leq \bar{\sigma}_V$ which with (5) yield

$$\sigma_N^2 = \frac{1}{N} \bar{\sigma}_V + \frac{N-1}{N} \bar{\sigma}_{CV} \leq \left[\frac{1}{N} + \frac{N-1}{N}\right] \bar{\sigma}_V = \bar{\sigma}_V. \tag{42}$$

This means that the error variance of the combined forecast is always smaller than or equal to the average error variance of all individual forecasts. From the intersection between $\bar{\sigma}_V$ and $\sigma_{N,max}^2(\bar{\rho})$ (see Figure 1), we finally find a $\bar{\rho}_x$ with

$$\bar{\rho}_x = \frac{1}{N-1} \left[\frac{\bar{\sigma}_V}{\sigma_{max}^2} - 1\right]. \tag{43}$$

Then, it is

$$\begin{aligned} \sigma_N^2 \leq \sigma_{N,max}^2 \leq \bar{\sigma}_V, \qquad & \text{if } \bar{\rho} \leq \bar{\rho}_x, \\ \sigma_N^2 \leq \bar{\sigma}_V \leq \sigma_{N,max}^2, \qquad & \text{if } \bar{\rho} > \bar{\rho}_x. \end{aligned} \tag{44}$$

ACKNOWLEDGEMENTS

This work has been realized in the framework of the public research project FOCCSI (Forecast Optimization by Correction and Combination methods for System Integration; 03ET4056) funded by the German Federal Ministry for Economic Affairs and Energy.

REFERENCES

Bates, J. M. and Granger, C. W. J. (1969) Combination of forecasts, Operational Research Quarterly, vol. 20, pp. 451-468.

Chan, F. and Pauwels, L. L. (2018) Some theoretical results on forecast combinations, International Journal of Forecasting, vol. 34, pp. 64-74.

Claeskens, G., Magnus, J. R., Vasnev, A. L. and Wang, W. (2016) The forecast combination puzzle: A simple theoretical explanation, International Journal of Forecasting, vol. 32, pp. 754-762.

Clemen, R. T. and Winkler, R. (1986) Combining economic forecasts, Journal of Business & Economic Statistics, vol. 4, pp. 39-46.

Clemen, R. T. (1989) Combining Forecasts: A Review and Annotated Bibliography, International Journal of Forecasting, vol. 5, pp. 559-581.

Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy, Journal of

Business & Economic Statistics, vol. 13, pp. 134-144.

Elliott, G. (2011) Averaging and the Optimal Combination of Forecasts, University of California, San Diego, USCD working paper.

Elliott, G. and Timmermann, A. (2016) Forecasting in Economics and Finance, Annual Review of Economics, vol. 8, pp. 81-110.

Genre, V., Kenny, G., Meyler, A., and Timmermann, A. (2013) Combining expert forecasts: Can anything beat the simple average? International Journal of Forecasting, vol. 29, pp.108-121.

Granger, C. W. J. and Ramanathan, R. (1984) Improved methods of combining forecasts, Journal of Forecasting, vol. 3, pp. 197-204.

Granger, C. W. J. and Jeon, Y. (2004) Thick modelling, Economic Modelling, vol. 21, pp. 323-343.

Guidolin, M. and Timmermann, A. (2009) Forecasts of US short-term interest rates: a flexible forecast combination approach, Journal of Econometrics, vol. 150, pp. 297-311.

Hendry, D. and Clements, M. (2004) Pooling of forecasts, The Econometrics Journal, vol. 1, pp. 1-31.

Jackson, T. and Karlsson, S. (2004) Finding Good Predictors for Inflation: A Bayesian Model Averaging Approach, Journal of Forecasting, vol. 23, pp. 479-498.

Madigan, D. and Raftery, A. (1994) Model selection and accounting for model uncertainty in graphical models using occam's window, Journal of the American Statistical Association, vol. 89, pp. 1335-1346.

Magnus, J. R. and De Luca, G. (2016) Weighted-average least squares (WALS): A survey, Journal of Economic Surveys, vol. 30, 117-148.

Makridakis, S. and Hibon, M. (2000) The M3-Competition: Results, Conclusions and Implications, International Journal of Forecasting, vol. 16, pp. 451-476.

Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2020) The M4 Competition: 100,000 time series and 61 forecasting methods, International Journal of Forecasting, vol. 36, pp. 54-74.

Marcellino, M. (2004) Forecast Pooling for Short Time Series of Macroeconomic Variables, Oxford Bulletin of Economic and Statistics, vol. 66, pp. 91-112.

Moral-Benito, E. (2015) Model Averaging in Economics: An Overview, Journal of Economic Surveys, vol. 29(1), pp. 46-75.

Smith, J. and Wallis, K. F. (2009) A Simple Explanation of the Forecast Combination Puzzle, Oxford Bulletin of Economics and Statistics, vol. 71(3), pp. 331-355.

Stock, J. H. and Watson, M. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, in *Festschrift in Honour of Clive Granger*, R. F. Engle and H. White Eds., pp. 1-44, 2001.

Stock, J. H. and Watson, M. (2004) Combination Forecasts of Output Growth in a Seven-Country Data Set, Journal of Forecasting, vol. 23, pp. 405-430.

Timmermann, A. Forecast combinations, in *Handbook of Economic Forecasting*, G. Elliott, C. W. J. Granger, and A. Timmermann Eds., North-Holland, 2006.