Reducing Estimation Errors in Forecast Combination: Covariance Matrix Estimation and Trimming Methods

Yannick van Etten

(2688877)

27 June 2023

Bachelor Thesis Econometrics and Operations Research

Thesis commission:

Dr. L. Hoesch (supervisor)   Xia Zou (co-reader)

Abstract

This thesis examines possible improvements in the estimation of forecast combinations. The aim is to minimize the estimation errors and to create a more effective combination. The trimming method explored by Radchenko et al. (2023) is investigated in detail since this diminishes the impact of estimation errors. In addition, this thesis focuses on minimizing estimation errors in the covariance matrix estimation process. Alternative estimation methods, such as linear shrinkage and factor models, are explored. The findings in this thesis show a remarkable improvement in the performance of forecast combinations when these methods are applied. The use of the linear shrinkage method can result in an improvement of up to 20% for GDP growth and more than 40% for unemployment forecasts compared to the equal weights method. The factor model method result in an improvements of up to 10% for GDP growth and more than 30% for inflation forecasts compared to the equal weights method.

Keywords: forecast combination, estimation errors, negative weights, covariance matrix, linear shrinkage, factor model, forecast combination puzzle.

1 Introduction

Accurate forecasts are critical in informed decision-making in today's rapidly changing and uncertain world. Recent global challenges, including economic shocks and the growing impact of climate change, have shown the importance of accurate forecasting. This raises a fundamental question: Is relying on a single forecast optimal, or can combining multiple forecasts lead to better results? This question has been extensively studied. Research by Bates and Granger (1969) shows that combining multiple individual forecasts leads, in theory, to a more accurate forecast. However, studies, such as Smith and Wallis (2009), have revealed situations where the estimated optimal forecast combination often fails to outperform the simple equal weights method calculated by the average of the individual forecasts. This difference between theoretical expectations and empirical outcomes is called the 'forecast combination puzzle'. This thesis explores this puzzle and seeks to improve the forecast combination by addressing the challenges of estimation errors, particularly in the covariance matrix estimation process.

Various approaches have been explored to tackle the issue of estimation errors in forecast combinations. A study by Radchenko et al. (2023) focuses on negative weights in forecast combinations. This factor is often ignored or trimmed to zero due to its less intuitive nature. Radchenko et al. (2023) demonstrate that negative weights can provide valuable information, particularly in highly correlated settings. The researchers propose a trimming method that addresses the forecast combination puzzle and captures the valuable information of negative weights. The variance is reduced by trimming the estimated optimal weights below a certain, potentially negative threshold, although at the cost of introducing some bias. This trimming method often outperforms the equal weights method and the untrimmed forecast combination as it diminishes the impact of estimation errors. Nevertheless, a fundamental question arises: Can the estimation process be enhanced to minimize the errors at their origin instead of merely diminishing the effect of estimation errors? Will an improved forecast combination be achieved by doing this?

This thesis aims to tackle these questions by focusing on methods that minimize estimation errors themselves, particularly in the estimation process of the covariance matrix, a critical step highlighted by Ledoit and Wolf (2004) and James and Stein (1961). This thesis first explores the effects of the trimming methods investigated in Radchenko et al. (2023). In addition, alternative approaches to covariance matrix estimation are explored, precisely the linear shrinkage method and the factor model. Through a simulation study involving multiple forecasts, the effectiveness of these methods is evaluated. Subsequently, these methods are applied to the empirical application conducted by Radchenko et al. (2023), comparing the performance of forecast combinations based on alternative covariance matrices with the equal weights method and the proposed trimming methods. The findings of this thesis demonstrate significant improvements in forecast combination performance, emphasizing the importance of reducing estimation errors.

Related Literature. This thesis relates to several pieces of literature. The theoretical analysis of forecast combination dates back to the works of Bates and Granger (1969) and Newbold and Granger (1974), which demonstrated the theoretical benefits of the combination compared to the individual forecasts and the equal weights forecasts in a fixed setting where the covariance matrix did not have to be estimated. However, studies, like Smith and Wallis (2009), have shown the difference between the theoretical and empirical performance of forecast combinations, the forecast combination puzzle. Claeskens et al. (2016) addressed this puzzle by explicitly taking into account the estimation step and its impact on the performance of forecast combinations, inspired by literature regarding model averaging. That paper finds that estimating the weights of the combination results in the presence of estimation errors. These errors increase the variance of the forecast combination. This leads to a decrease in the functionality of the optimal forecast combination compared to the equal weights method. This thesis builds upon these findings and

explores methods to reduce these estimation errors.
Radchenko et al. (2023) investigated the presence of negative weights in forecast combinations. Traditionally, negative weights are ignored or trimmed to zero due to their less intuitive nature. However, the paper showed that negative weights can provide valuable information, particularly in positive highly correlated settings. The researchers propose a trimming method with a negative threshold that addresses the forecast combination puzzle and captures the valuable information contained in negative weights. This method reduces the variance of the weights and results in a forecast combination that often outperforms the equal weights method and the forecast combination based on untrimmed weights. This work creates awareness of the usefulness of negative weights. In this thesis, negative weights will not be ignored. The simulation study and empirical application in Radchenko et al. (2023) provide a suitable setting to evaluate forecast combinations based on different covariance matrix estimates. This thesis will explore the performance of the forecast combinations based on the linear shrinkage method and the factor method, where the sample covariance matrix is used in Radchenko et al. (2023).
Ledoit and Wolf (2004)), Ledoit and Wolf (2021) and James and Stein (1961)) highlight the challenges associated with estimating covariance matrices when the number of parameters is comparable to the number of observations. These papers propose the linear shrinkage method, which combines the sample covariance matrix with a structured target matrix to achieve more accurate estimates of the covariance matrix. Ross (1976) and Fan, Liao, et al. (2013) propose the factor model as an alternative covariance estimation method. While these covariance matrix estimation methods are widely used in portfolio theory, their application to forecast combination, particularly when negative weights are considered, remains limited. This thesis will contribute to the existing literature by applying and evaluating these methods in a forecast combination setting.
The remainder of this thesis is structured as follows. Section 2 discussed the theory behind forecast combination and the presence of negative weights. Section 3 presents the theory behind methods that could improve the forecast combination. Different trimming methods and covariance estimation methods will be explored. Section 4 will evaluate the trimming method in a simulation study and an empirical application, thereby replicating Radchenko et al. (2023). Section 5 investigates the impact of different covariance matrix estimation methods on the forecast combination performance in a simulation study and in the empirical application. Section 6 concludes. The Python code for the conducted simulation studies and the specific R code is publicly available.[1]

[1]The Python code for the simulation studies is available on: https://github.com/YannickvanEtten/bachelor_thesis. Two Jupyter Notebooks and an R file are available. The first Notebook shows a detailed replication of the simulation study conducted by Radchenko et al. (2023). The second Notebook illustrates the impact of covariance estimation on the forecast combination. In both Notebooks, the methods, theory, and functions are explained in detail. The R code for the data-driven threshold method with correct p-values are available, solving a mistake by Radchenko et al. (2023).

2 Theoretical Analysis of Forecast Combination

In many situations, decision-makers are presented with multiple forecasts for the same variables. For example, the European Central Bank (ECB) Survey of Professional Forecasters (SPF) provides quarterly forecasts for inflation, real Gross National Product (RGDP), and unemployment, involving around 100 advisors.[2] When faced with multiple forecasts, decision-makers need to determine how to use these forecasts.

This section provides a theoretical analysis of forecast combination. The intuition behind combining forecasts is explored. Additionally, the challenges associated with estimating the weights for these combinations are discussed. The concept of negative weights will also be explored and their potential relevance in specific settings.

2.1 Intuition behind Forecast Combination

The objective of this subsection is to gain a comprehensive understanding of the challenges associated with handling diverse forecasts. In J. R. Magnus and Luca (2016) the concept of model averaging is discussed. In the paper, the intuition of model averaging is illustrated by a fictional story about a king. An adaption of this story illustrates the concept of forecast combination. In this story a king is seeking to forecast inflation for the next year and is consulting 12 advisors. The king has three options for dealing with these 12 different forecasts. First, he could evaluate each advisor based on certain criteria and select the forecast of the most trusted or best-performing advisor. Alternatively, he could treat all advisors equally and take the average of their forecasts. However, the king may have prior knowledge or preferences regarding the advisors' expertise, leading to a third method where more competent advisors receive higher weights.

The question arises: which method works best? In standard econometric modeling, the first method of selecting the best-performing model is often favored, based on various performance indicators as is seen in J. Magnus (1999). However, the third method, which considers all perspectives of the advisors, is more intuitive. Nevertheless, the second method, where all advisors are trusted equally, is commonly employed in practice, as seen in Claeskens et al. (2016). This is mainly due to the difficulties in estimating optimal weights for the second method. Claeskens et al. (2016) finds that estimation errors are common. To avoid this, all weights are often set equal. This modified story from J. R. Magnus and Luca (2016) is used in this thesis because of its simplicity and clear illustration of the different methods for dealing with forecasts. Furthermore, this story will be used to clearly illustrate the use of negative weights in Subsection 2.4.

2.2 Forecast Combination in a Theoretical Setting

When multiple advisors provide forecasts, each advisor may use different models, assumptions, and information sets, leading to variations in their predictions. It is uncertain whether a single advisor consistently outperforms others, leading to the concept of forecast combination. The idea is to use all the valuable information from each independent advisor to create a combined forecast that outperforms all individual advisors. One of the first papers exploring the potential benefits of forecast combination is Bates and Granger (1969). Their intuition is that every forecast has

[2] The European Central Bank (ECB) Survey of Professional Forecasters (SPF) provides quarterly forecasts for inflation, real Gross National Product (RGDP), and unemployment of different horizons. The earliest available survey data is from Q3 1999. In the empirical application of this thesis, the data of the one- and two-year-ahead forecast horizons for inflation, RGDP and unemployment are used until Q2 2018. The data from ECB SPF is publicly available on: http://www.ecb.europa.eu/stats/prices/indic/forecast.

some type of useful independent information. The paper highlights two types of independent information:

- One forecast may be based on variables of information that the other forecasts did not consider.
- One advisor may have considered different assumptions about the functional form of the relationship between the variables.

Bates and Granger (1969) argue against removing forecasts with useful independent information. The initial combination method that comes to mind is equal weights, commonly used as the baseline approach. With the equal weights method, the combined forecast is calculated as the average of the individual forecasts. This method avoids the estimation errors caused by weight estimation and it has the potential to outperform individual forecasts. The question remains whether it is possible to determine these weights intelligently, taking into account their past reliability. Forecast methods are often evaluated using the mean-squared error (MSE). Various methods have been developed to determine weights and create a forecast combination with the lowest MSE, moving away from the equal weights method

In Bates and Granger (1969) the forecast combination problem is first explored in a theoretical setting to estimate an event with a true constant value μ. It assumes known constant variances σ_1^2 and σ_2^2 for the two forecasts over time t. The two forecasts, denoted as $y_{1,t}$ and $y_{2,t}$, are unbiased, with $E(y_{1,t}) = E(y_{2,t}) = \mu$. It is important to note, that in a setting where the value μ is changing over time. The variance and covariance terms are determined by the forecast error, instead of the forecast itself. The setting of a time-varying μ is frequently used, in sections of this thesis as well. By combining the forecasts using weights w and $w - 1$, a forecast combination $y_{c,t}$, is created:

$$y_{c,t} = wy_{1,t} + (w-1)y_{2,t}$$

This forecast combination is unbiased as well. The variance of the forecast combination is determined as:

$$\sigma_c^2 = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2\rho w\sigma_1(1-w)\sigma_2$$

where $\rho = \text{corr}(y_1, y_2)$ represents the correlation between the first and the second forecast. To obtain the smallest MSE value, the variance of the forecast combination σ_c^2 is minimized. Taking the derivative of σ_c^2 with respect to the weights and setting this equal to zero results in the optimal weight:

$$w^* = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \tag{1}$$

Using the optimal weight, the expression for the optimal forecast combination can be determined as $y_c^* = w^*y_1 + (1-w^*)y_2$. The variance of y^* is given by:

$$\text{var}(y^*) = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \tag{2}$$

If an individual forecast alone would be best-performing, the optimal weight w^* will be 0 or 1. This is only the case when $\sigma_1 = \sigma_2$ and $\rho = 1$ or in the case $\rho = \sigma_1/\sigma_2$. This results in an equal performance of the combination and the individual forecast. In all other cases, the variance of the combination is lower than the variances of both individual forecasts.

The baseline method in forecast combination is equal weights, $w = 0.5$, resulting in a forecast combination of $y_{eq} = 0.5y_1 + 0.5y_2$. This combination has a variance of:

$$\text{var}(y_{eq}) = \frac{1}{4}\sigma_1^2 + \frac{1}{4}\sigma_2^2 + \frac{1}{2}\sigma_1\sigma_2\rho \tag{3}$$

Based on these expressions (2) and (3), equal weights are only optimal when $\sigma_1^2 = \sigma_2^2$. In all other cases, the variance of the forecast combination using optimal weights is smaller than the variance of equal weights. This indicates that combining forecasts with optimal weights can be beneficial. In Newbold and Granger (1974), this concept is extended to n unbiased forecasts. The forecast combination is rewritten as:

$$y_c = \mathbf{w}'\mathbf{y}$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$ and forecast combination y_c is unbiased. This combination has a variance of:

$$\text{var}(y_c) = \mathbf{w}'\Sigma\mathbf{w}$$

where Σ represents the true covariance matrix of the forecast error. The variance of the forecast combination will be minimized, resulting in the optimal weights vector:

$$\mathbf{w}^* = \frac{\Sigma^{-1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \tag{4}$$

where $\mathbf{1}$ represents a vector of ones. This equation will also be used in settings where the covariance matrix is unknown and estimated by $\widehat{\Sigma}$.

This analysis highlights that the performance of forecast combinations depends on the quality of the individual forecasts. Until now, an optimal linear combination of forecasts in terms of the MSE is explored in a setting where the covariance matrix is assumed to be known. In this setting, the use of forecast combination leads to an improved forecast compared to all individual forecasts and the equal weights method. In practice, the variance and the covariance are unknown. The covariance matrix must be estimated, leading to estimated optimal weights and an estimated optimal forecast combination. Therefore, the performance of the combination also relies on the accuracy of the weight estimation process. This will be explored in the next subsection.

2.3 Weight Estimation in Forecast Combination

The forecast combination using the optimal weights w^* results in an improved forecast in a theoretical setting. In this subsection, the estimation aspects will be introduced. In particular, the covariance matrix needs to be estimated in order to obtain estimated optimal weights $\widehat{w}$. The estimation of weights in forecast combination introduces estimation errors, a reevaluation of the optimality of the estimated optimal forecast combination is necessary. In this setting, the forecast combinations are evaluated by the mean squares forecasting error (MSFE). Empirical studies and simulations, such as Smith and Wallis (2009), have shown that the estimated optimal forecast combination often does not perform as well as expected in theory and may not outperform the equal weights method. This has been called the 'forecast combination puzzle' by Smith and Wallis (2009) and could be due to the presence of estimation errors.

The estimation errors are studied in detail in the paper of Claeskens et al. (2016). The estimation step is taken into account explicitly, inspired by literature regarding model-averaging. The paper extends the forecast setting of Bates and Granger (1969) by estimating the optimal weights. The optimal weights become random by the estimation part rather than fixed. This leads to a forecast combination using estimated weight $\widehat{w}$:

$$y_c = \widehat{w}y_1 + (1 - \widehat{w})y_2$$

When the forecasts and weights (y_1, y_2, w) follow a trivariate normal distribution, the forecast combination becomes biased even if the original forecasts are unbiased.

$$E(y_c) = \mu + \text{cov}(\widehat{w}, y_1 - y_2)$$

The variance is in this case determined by:

$$\begin{aligned}\text{var}(y_c) =& E(\widehat{w})^2\sigma_1^2 + (1-E(\widehat{w}))^2\sigma_2^2 + 2\rho E(\widehat{w})\sigma_1(1-E(\widehat{w}))\sigma_2 \\ &+ \text{var}(\widehat{w})\text{var}(y_1-y_2) + (\text{cov}(\widehat{w}, y_1-y_2))^2\end{aligned} \tag{5}$$

In the case where the weights are independent of (y_1, y_2), the forecast combination remains unbiased and the variance is given by:

$$\begin{aligned}\text{var}(y_c) = & E(\widehat{w})^2\sigma_1^2 + (1-E(\widehat{w}))^2\sigma_2^2 + 2\rho E(\widehat{w})\sigma_1(1-E(\widehat{w}))\sigma_2 \\ &+ \text{var}(\widehat{w})\text{var}(y_1-y_2)\end{aligned} \tag{6}$$

In equations (5) and (6) the variance of y_c increases compared to the setting with fixed weights. The variance of the fixed setting is seen in the first three terms of (5) and (6), however, an additional positive component is present. This increase in variance is an additional component introduced by the weight estimation process.

This finding implies that weight estimation introduces more uncertainty into the forecast combination, leading to an increase in the variance of the combination. Consequently, using estimated optimal weights may not always outperform equal weights anymore. However, there are methods available to reduce the variance of the combination, examined in Section 3. One approach is the use of trimming methods, this reduces the variance while introducing some bias. This method is discussed in Subsection 3.1. Another method is the use of a different covariance estimation techniques. While the sample covariance estimation method is commonly used in practice, as seen in Radchenko et al. (2023), it can lead to estimation errors when the number of forecasts is close to the number of observations. Using an alternative covariance estimation method can result in a lower variance of the weights and a more effective forecast combination. These methods will be discussed in Subsection 3.2. The concept of negative weights and their potential applications are explored first.

2.4 Negative Weights in Forecast Combination

In the previous part, an expression for the optimal weights in the forecast combination is derived. Given specific variance and covariance values the optimal weight could become negative. In this subsection, the conditions for negative weights will be discussed. Furthermore, the effect of the negative weights will be illustrated.

In the setting of Bates and Granger (1969) with two forecasters where the variance is assumed to be known, the expression for the optimal weights is derived in Equation (1). To illustrate the effects, this is repeated:

$$w^* = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

Derived from this expression, the optimal weight w^* is negative when $\rho < \sigma_2/\sigma_1$. This indicates that negative weights are optimal when the correlation between the forecasts is large and positive, derived in Radchenko et al. (2023) as well. When the correlation is zero or negative, the weights will be positive.

Positive weights in forecast combinations are intuitive as they assign a percentage to each forecast based on its reliability (smaller variance), ensuring that the forecast combination falls within the range of the lowest and highest forecasts. However, a large and positive correlation between the forecasts,$\rho < \sigma_2/\sigma_1$, is not an inequality that is hard to fulfil. Making negative optimal weights not uncommon. What is the intuition behind these negative weights?

An adaption of the story of the king is seen in Chapter 5 in J. R. Magnus and Luca (2016). This illustrates the effect of the negative weights. In this scenario, the king wants a forecast for next year's inflation from two advisors. The first advisor forecasts 2% inflation with a variance of 1, while the second advisor forecasts 4% inflation with a variance of 4. If these advisors are uncorrelated, meaning they have access to different information sets and have not communicated with each other, the forecasts will be weighted using positive weights. Since the first advisor has a smaller variance, it will receive a higher weight than the second advisor. Therefore, the forecast combination will be closer to the first forecast, resulting in an inflation forecast of 2.5%.

However, the assumption of no correlation between the advisors is unrealistic. It is likely that their information sets overlap, and they often work together, resulting in a positive correlation; a correlation of 3/4 is assumed. This would result in a forecast combination of 1.5%, using Equation (1). The correlation resulted in the optimality of a negative weight. Consequently, the combination is now outside the 2% and 4% interval.

From a theoretical standpoint, the presence of negative weights in forecast combinations makes sense. Both advisors have similar information; however, the second advisor is less certain due to the high variance and, therefore, further away from the true value compared to the first advisor. This is an indication that both forecasts overestimate inflation. While the first advisor is the most trustworthy, the second advisor reveals information about how much both overestimate the true inflation. With the negative weight, this overestimation can be taken into account, as is seen in Radchenko et al. (2023) as well.

However, in practical situations, the use of negative weights will raise questions for those who are not familiar with the technical aspects of forecast combinations. It can be challenging to explain why the combination falls outside the range of individual forecasts, especially when a decision-maker, such as the king, decides to use a different forecast estimate from both advisors. Due to these challenges, negative weights are often set to zero in practice, as is observed in J. R. Magnus and Luca (2016). It is important to note that this introduces biases and may lead to suboptimal forecast combinations.

Given an understanding of the impact of negative weights, the prevalence of negative weights in a setting with more than two forecasters is explored, following the paper of Radchenko et al. (2023). Of interest is the forecast combination $y_c = \mathbf{w}'\mathbf{y}$, where the vector of forecasts is given by $\mathbf{y} = (y_1, \ldots, y_n)'$. The optimal weight vector $\mathbf{w}^*$ for these individual forecasts is obtained in Equation (4). This weight vector can be partitioned into two parts, representing negative and positive weights. The resulting vector can be expressed as:

$$\mathbf{w}^* = \begin{bmatrix} \mathbf{w}_-^* \\ \mathbf{w}_+^* \end{bmatrix}$$

Similarly, the covariance matrix and the unit vector can be partitioned in a corresponding manner:

$$\Sigma = \begin{bmatrix} \Sigma_{--} & \Sigma_{-+} \\ \Sigma_{+-} & \Sigma_{++} \end{bmatrix} \qquad \mathbf{1} = \begin{bmatrix} \mathbf{1}_- \\ \mathbf{1}_+ \end{bmatrix}$$

The inversion formula for the block matrix Σ is used to derive the following result:

$$w_-^* = \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}(E^{-1}\mathbf{1}_- - E^{-1}\Sigma_{-+}\Sigma_{--}^{-1}\mathbf{1}_+)$$

where $E = \Sigma_{--} - \Sigma_{-+}\Sigma_{++}^{-1}\Sigma_{+-}$ represents the Schur complement. This result leads to Proposition 1, as derived in Radchenko et al. (2023).

Proposition 1. *The vector of optimal weights contains negative elements $w_-^* < 0$ is and only if $E^{-1}\mathbf{1}_- < E^{-1}\Sigma_{-+}\Sigma_{--}^{-1}\mathbf{1}_+$ element wise.*

Large positive elements in Σ_{-+} increase the likelihood of negative weights. This indicates that negative weights tend to appear when the available individual forecasts are highly positively correlated, similar to the setting with two forecasts. This situation may arise when all forecasts consistently overestimate or underestimate the true value. Negative weights are introduced to correct

this, allowing the combined forecast to be outside the range of the individual forecasts.
Although the concept of negative weights is less intuitive than positive weights, their presence is of importance in the theoretical setting. In practice the weights are often set at zero, this may lead to an increase in bias. However, setting these weights to zero can reduce the variance of the weights. The trade-off between the increase in bias and the decrease in variance will be further discussed in Subsection 3.1.

2.5 Conclusion Theoretical Analysis

This subsection shows that the optimal combination of all forecasts leads to an improved forecast. In this theoretical setting, the covariance matrix is assumed to be known, and the combination is compared to the equal weights method. In practice, the covariance matrix must be estimated. The weight estimation introduces more uncertainty into the forecast combination, leading to an increase in its variance. Consequently, using estimated optimal weights may not consistently outperform equal weights anymore. Alternative methods can be used to increase the effectiveness of the estimated optimal forecast combination.
Furthermore, while negative weights are less intuitive than positive weights, their presence is important. In this section, it was derived that negative weights appear when the available individual forecasts are highly positively correlated, for example, when all forecasts consistently overestimate or underestimate the true value. Negative weights can correct this, allowing the combined forecast to be outside the range of the individual forecasts. In practice, the weights are often set at zero. This may lead to an increase in bias.

3 Performance-improving Techniques for Forecast Combination

In this section two different ways for improving optimal weight estimation are theoretically evaluated: different trimming methods and alternative covariance estimation methods.

3.1 Trimming of Weights

The trimming method is commonly used to address the increased variance of the forecast combination. This method is frequently applied to use all weights in the range of $[0, 1]$, as seen in Smith and Wallis (2009). The trimming method will reduce the variance of the weights while it will also increase the bias. The nonzero trimming method is defined as:

$$w^{\text{TR}} = \max(\widehat{w}, 0) = \begin{cases} \widehat{w}, & \widehat{w} \geq 0 \\ 0, & \widehat{w} < 0 \end{cases} \tag{7}$$

The trimming method introduces a bias-variance trade-off. A similar trade-off happens in the equal-weight method, discussed earlier. Depending on the situation, this trade-off can be beneficial. The choice to trim at zero is arbitrary. In Subsection 2.4, it was concluded that negative weights could have an important function in theory. This suggests that it could be beneficial to trim at a different negative value to achieve a better balance between variance and bias, this has been explored in Radchenko et al. (2023). The trimming methods in this paper will be used in the thesis as well.
There are two types of trimming methods that ensure all weights are above a threshold $c \leq 0$. The first is a two-step trimming method, where the optimal weights are calculated first and then trimmed if they are below the threshold. The second is a one-step trimming method, where the

weights are estimated and simultaneously trimmed using constrained optimization. These methods, which apply the same threshold to all n weights, will be further discussed in this subsection. Additionally, a data-driven threshold method will also be discussed. Where the threshold will be determined based on the available data.

3.1.1 Two-step Trimming Methods

The first trimming method involves truncating the weights that are below a given threshold, denoted as c. This method is similar to the zero trimming method (7), but c replaces zero as the threshold. For estimated weights, $\widehat{w} = (\widehat{w}_1, \ldots, \widehat{w}_n)'$, the method is defined as follows:

$$w_i^{\text{TR1}} = \alpha_1 \times \begin{cases} \widehat{w}_i, & \widehat{w_i} > c \\ c, & \widehat{w}_i \leq c \end{cases} \tag{8}$$

where α_1 is a scaling parameter introduced to ensure that the trimmed weights still sum up to one. As a result of this rescaling, the weights that were below the threshold will be adjusted again. The new vector of trimmed weights will be biased, but its variance will be lower than that of the vector of estimated untrimmed weights.
Alternatively, it is possible to rescale only the weights that are not trimmed. This approach is employed in the second trimming method:

$$w_i^{\text{TR2}} = \times \begin{cases} \alpha_2 \times \widehat{w}_i, & \widehat{w_i} > c \\ c, & \widehat{w}_i \leq c \end{cases} \tag{9}$$

Again, α_2 is a rescaling parameter that ensures the trimmed weights sum up to one. In both methods (8) and (9), all weights below the threshold are treated in the same way, without considering the magnitude of the negativity. However, the magnitude of negativity can provide valuable information. This information can be incorporated into the trimming process, which is done in the third trimming method:

$$w_i^{\text{TR3}} = \times \begin{cases} \alpha_3 \times \widehat{w}_i, & \widehat{w_i} > c \\ \dfrac{\widehat{w}_i}{\text{min } \widehat{w}} \times c, & \widehat{w}_i \leq c \end{cases} \tag{10}$$

This method can be interpreted as follows: the most negative weight below the threshold c will be set to c. All remaining weights below the threshold will be transformed to a value on the $[c, 0]$ interval based on the relative negativity captured by the fraction of $\frac{\widehat{w}_i}{\text{min } \widehat{w}}$. As in the previous methods, α_3 is included to ensure that the trimmed weights still sum up to one in method (10) as well.

3.1.2 One-step Trimming Methods

In the one-step methods, the estimation and trimming of the weights are conducted through constrained optimization, following Radchenko et al. (2023). The objective of the optimization is to minimize the variance of the forecast combination $w'\Sigma w$ with a constraint on the weight values, which is incorporated into the optimization process. This leads to the fourth trimming method:

$$\begin{aligned} w^{\text{TR4}} = \text{arg } \min_{c} \; & w'\Sigma w \\ & w'\mathbf{1} = 1 \\ & w_i \geq c \end{aligned} \tag{11}$$

The last method introduces an L1-norm constraint in the optimization problem. Instead of directly restricting the weights, the L1-norm of the weights, given by $||w||_1 = \sum_{i=1}^{n} |w_i|$, is constrained.

This approach follows a recommendation from Fan, Zhang, et al. (2012) and leads to the fifth trimming method:

$$\begin{aligned} w^{\text{TR5}} = \arg\min_{c}\, & w'\Sigma w \\ & w'\mathbf{1} = 1 \\ & ||w||_1 \leq 1 - \tilde{c} \end{aligned} \tag{12}$$

Setting $\tilde{c}$ to zero imposes a nonnegativity constraint on the weights. When $\tilde{c}$ is set to $+\infty$, the constraint is ignored, and no trimming is applied. For values of $0 < \tilde{c} < +\infty$, the weights are allowed to be negative but bounded. In this case, the weights are not directly bounded by $\tilde{c}$ itself, but rather by a number associated with $\tilde{c}$. Unlike method (11), the L1-constraint in method (12) puts an upper bound on the weights as well.

3.1.3 Data-driven Threshold Method

Depending on the situation, the trimming methods could have a positive effect on the forecast combination. The question remains how to set the threshold parameter c since this influences the bias-variance trade-off. This could be specified by the researchers. A more intuitive method is also possible. The aim of this method is to let the data determine the optimal choice of the threshold. The threshold will be based on the preudo-out-of-sample MSFE, following Radchenko et al. (2023). Given a time series of length T, this data set will be split into two parts, $[1, t_k - 1]$ and $[t_k, T]$ where $t_k < T$ is a time index. The first part of length $t_k - 1$ is used to estimate the covariance matrix and the weight vector $\widehat{w}$. The second part is to determine the MSFE value of the specified trimming method TR and threshold value c. This part has a length of $T - t_k$ as the last observation T is used only to calculate the forecasting error. The MSFE is calculated as follows:

$$\text{MSFE}(c, t_k) = \frac{1}{T - t_k + 1} \sum_{t=t_k}^{T-1} (\mathbf{y}_t'\mathbf{w}^{\text{TR}}(c) - \mu_{t+1})^2 \tag{13}$$

where $\mathbf{y}_t' = (y_{1,t}, \ldots, y_{n,t})$ is the vector of individual forecasts for a variable μ at time $t+1$. The true value is indicated by μ_{t+1}. The forecast error of the forecast combination using a specific trimming method and a threshold is determined by $\mathbf{y}_t'\mathbf{w}^{\text{TR}}(c) - \mu_{t+1}$.

To increase the stability and reliability of the MSFE seen in Equation (13), the average MSFE (AMSFE) is considered, following Fan, Zhang, et al. (2012) and Radchenko et al. (2023). To determine the AMSFE, the data set is partitioned in several ways, by varying the partition parameter t_k. Therefore the AMSFE given a specific threshold c and value K, is derived by:

$$\text{AMSFE}(c) = \frac{1}{K} \sum_{t_k=1}^{K} \text{MSFE}(c, t_k) \tag{14}$$

The aim is to determine the threshold c that minimizes the AMSFE. In this way, c^* is determined by:

$$c^* = \arg\min_{c} \text{AMSFE}(c) \tag{15}$$

This results in optimal threshold c^*. Using the optimal threshold, the forecast combination, defined by: $\mathbf{y}_T'\mathbf{w}^{\text{TR}}(c^*)$ for variable μ_{T+1} with a data-driven threshold method.

3.2 Covariance Estimation Methods

In the last subsection, the trimming method was used to decrease the variance of the weights caused by estimation errors. However, it could also be beneficial to improve the way the weights are estimated. The weights are often determined by the use of the covariance matrix of the forecast error. This subsection focuses on analyzing alternative methods for estimating the covariance matrix. This begins by addressing the potential issues known to covariance estimation in high-dimensional settings. In particular, the effect on the sample covariance estimation method, since this method is often used in practice. Subsequently, various alternative covariance matrix estimation methods are evaluated.
As alternatives the linear shrinkage method and the factor model approach will be explored. These approaches have been applied in various fields, including finance and in particular portfolio optimization, where high-dimensional covariance matrices are common too. Research studies by Fan, Liao, et al. (2013) and Ledoit and Wolf (2021) have demonstrated the effectiveness of these methods in improving estimation in portfolio optimization. Making these two methods suitable candidates for the forecast combination setting.

3.2.1 Covariance Estimation Setting

This part aims to evaluate potential covariance estimation problems that can arise in a setting where the number of forecasters is close to the number of available observations, referred to as a high-dimensional setting. The problem is called the 'curse of dimensionality', and is often present in forecast combination settings.
Let p denote the number of forecasters and n represent the number of observations. This leads to a matrix with dimensions of $p \times n$ where each forecaster's forecast errors are recorded over time. The covariance matrix, Σ, has dimensions of $p \times p$, as it captures the relationships among the forecasters. The number of parameters, denoted as k, that need to be estimated for the covariance matrix is given by $k = p * (p + 1) * \frac{1}{2}$, taking into account the diagonal structure of the matrix. The number of variables to be estimated can cause problems

3.2.2 Sample Covariance Method

In practice, the sample covariance method is often used to estimate the covariance matrix. This method is used in the simulation study and the empirical application of Radchenko et al. (2023) as well. The sample covariance matrix estimate is denoted as $\widehat{\Sigma}_{sample}$. The sample covariance matrix is calculated by using the data matrix X_n, with dimensions $p \times n$:

$$\widehat{\Sigma}_{sample} = \frac{1}{n-1} X'X \tag{16}$$

The sample covariance method is unbiased and is the maximum likelihood estimator under normality, as is seen in Ledoit and Wolf (2004). However, it may not be the optimal estimator in high-dimensional settings. If $p > n$, the sample covariance matrix will have $p - n$ eigenvalues equal to zero. As a result, the matrix becomes semi-positive definite, singular, and non-invertible, as shown in Ledoit and Wolf (2021). On the other hand, if $p < n$ and the ratio of p/n is not negligible, the sample covariance estimate is invertible but numerically ill-conditioned, as highlighted in Ledoit and Wolf (2004). Consequently, inverting this matrix can lead to increased estimation errors. Only if the ratio of p/n is negligible small the sample covariance matrix is well defined and suited to use.

3.2.3 Linear Shrinkage Method

The method of linear shrinkage is an alternative approach for estimating the covariance matrix. The idea of linear shrinkage was first proposed by James and Stein (1961). The paper aimed to estimate the multivariate mean. It became clear that for dimensions greater than 3 ($p > 3$), there existed a better estimator than the sample mean. The sample mean is an estimator with similar unbiased properties as the sample covariance and is the maximum likelihood estimator. By combining the sample mean with a zeros vector, an estimator with a lower mean squared error (MSE) than the original sample mean could be obtained. This reduction in MSE was achieved by trading off some bias for a decrease in variance.

In the paper by James and Stein (1961), it became clear this insight into bias-variance trade-off was not merely useful in the case of the sample mean estimation. By combining the sample covariance matrix with a structured target matrix, such as the identity matrix, a lower MSE could be obtained compared to the sample covariance matrix alone. This method is called shrinking since the sample covariance matrix is shrunk to a structured target matrix.

Ledoit and Wolf (2004) extended the method of linear shrinkage. The idea behind their estimation method remains the same: finding a balance between the sample covariance matrix and a structured target matrix, such as the diagonal identity matrix. With this balanced shrinkage method, the goal is to reduce the impact of noise and increase stability, leading to a potentially more efficient estimator in high-dimensional data settings.

In the paper of Ledoit and Wolf (2004), the optimization of the difference between the true covariance matrix Σ and the shrinkage covariance matrix $\widehat{\Sigma}_{shrnk}$ is carried out using the Frobenius norm. The Frobenius norm of a square matrix A with dimensions $p \times p$ is defined as:

$$||A||_F = \sqrt{\langle A, A \rangle} = \sqrt{Tr(AA')/p} = \sqrt{\frac{1}{p} \sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij}^2} \tag{17}$$

where a_{ij} represents the element in the i-th row and j-th column of the matrix A. Using definition (17) the MSE is given by:

$$\text{MSE} = E[||\widehat{\Sigma}_{shrnk} - \Sigma||_F^2]$$

The aim is to minimize the MSE by constructing a shrinkage covariance matrix, $\widehat{\Sigma}_{shrnk}$, as a combination of the sample covariance matrix $\widehat{\Sigma}_{sample}$ and a structured target matrix. In Ledoit and Wolf (2004) the identity matrix is chosen. Resulting in the following optimization:

$$\begin{aligned} &\min_{\rho_1, \rho_2} E[||\widehat{\Sigma}_{shrnk} - \Sigma||_F^2] \\ &\text{s.t. } \widehat{\Sigma}_{shrnk} = \rho_1 I_p + \rho_2 \widehat{\Sigma}_{sample} \end{aligned}$$

A rewritten solution to this optimization problem is given by:

$$\widehat{\Sigma}_{shrnk} = \gamma_p^* \mu_p I_p + (1 - \gamma_p^*) \widehat{\Sigma}_{sample} \tag{18}$$

where $\gamma_p^* = \frac{\beta_p^2}{\delta_p^2}$ with $\beta_p^2 = E[||\widehat{\Sigma}_{sample} - \Sigma||_F^2]$, $\delta_p^2 = E[||\widehat{\Sigma}_{sample} - \mu_p I_p||_F^2]$ and $\mu_p = \langle \Sigma, I_p \rangle$. The equation can be interpreted as the sample covariance matrix is shrunk toward the target matrix $\mu_p I_p$ with shrinkage coefficient $\gamma_p^* \in [0, 1]$.

The resulting covariance estimate $\widehat{\Sigma}_{shrnk}$ is positive definite and invertible, even when $p > n$. It is important to note that the values μ_p, δ_p, and β_p rely on the true covariance matrix. Ledoit and Wolf (2004) provide a detailed estimation procedure for these quantities. Their paper is followed for the estimation of parameter μ_p, with $\widehat{\mu}_p = \langle \widehat{\Sigma}_{sample}, I_p \rangle$. This is equal to the average of the estimated variances of the sample covariance matrix for all forecasts.

Various shrinkage methods have been developed based on the same balancing idea between the sample covariance matrix and a well-structured matrix. The other methods differ in the choice of the well-structured matrix and the approach used to estimate the shrinkage parameter. In most forecast combination settings no pre-knowledge about the covariance is assumed, therefore the identity matrix is used as a structured target matrix in this thesis. Using a custom-tailored target matrix requires knowing features of the true covariance matrix structures, often available in finance settings, as is discussed in Ledoit and Wolf (2021).

Using only a scaled identity matrix as covariance matrix results in equal weights. This is a method often used in practice. Equal weights are also often used as a benchmark since the use of fixed weights reduces the estimation errors by setting the variance to zero while allowing for bias, as seen in Radchenko et al. (2023) as well. This creates an interesting setting where the shrinkage method balances the bias of the equal weights with the uncertainty of the sample covariance matrix.

As stated before, the paper of Ledoit and Wolf (2004) is followed for the estimation of parameter μ_p, with $\widehat{\mu}_p = \langle \widehat{\Sigma}_{sample}, I_p \rangle$ seen in Equation (18). These settings lead to the following linear shrinkage estimator used in this thesis, $\widehat{\Sigma}_{LS}$:

$$\widehat{\Sigma}_{LS} = \delta \widehat{\mu}_p I_p + (1 - \delta)\widehat{\Sigma}_{sample} \tag{19}$$

In this calculation, δ is defined as the shrinkage coefficient or the shrinkage intensity. If $\delta \approx 1$, it means that the identity matrix is fully utilized and the sample covariance matrix has minimal influence. On the other hand, if $\delta \approx 0$, the original sample covariance matrix is used without much influence of the identity matrix. Just as the weights are used to combine forecasts, the δ can be seen as a weight that combines these two matrices. There are different ways to determine the optimal value of δ. In this thesis, the δ value will be determined using a data-driven method, discussed in Subsection 3.1.3.

As mentioned before, after Ledoit and Wolf (2004) many adapted methods of shrinkage have been developed. This basic variant of the linear shrinkage method has been chosen in this paper for two reasons. First, most other methods use a custom-tailored target matrix, the use of this could potentially lead to improvements. As stated before, this requires knowledge of features of the true covariance matrix structures. Second, clear intuition behind balancing the equal weights method by the sample covariance method, making it possible to use beneficial characteristics from both models.

3.2.4 Factor Model Method

The last covariance estimation method that is explored is the factor model. This model is a widely utilized method in finance since the concept has been applied by Ross (1976) in the article concerning The Arbitrage Theory of Capital Asset Pricing. The factor model is used to predict and assess return and construct portfolios. The paper of Fan, Liao, et al. (2013) uses a factor model to estimate the covariance matrix in a high-dimension case, suggesting its potential use in determining the covariance matrix in the forecast combination setting.

In this context, subscript $i \in 1, \ldots, p$ represent the different forecasts, the subscript $t \in 1, \ldots, n$ denote the observations, the subscript $k \in 1, \ldots, K$ refers to the factors. Since, in the simulation and the empirical application the covariance matrix is determined by the forecast error, e_t, this is the focus here as well.

In the factor model framework, there exist different types of factors. There are observed factors and latent factors. Observed factors are known and can be based on external information or proposed by existing literature. A well-known example of observed factors is the Fama-French

factors. On the other hand, latent factors are unknown and need to be estimated from historical data using methods such as principal components or maximum likelihood estimation. In this paper, no pre-knowledge is available about the factors so latent factors are used.
Within the factor model framework, there are different types of models, static and dynamic models. This paper will focus on a static factor model. The static model assumes for every forecast i, the observed forecast error can be expressed as follows:

$$e_{i,t} = \beta_i' f_t + u_{i,t}$$

where the vector of factor loadings is denoted by $\beta_i = (\beta_{i_1}, \ldots, \beta_{i_K}\})'$, the vector of common factors represented by f_t a $K \times 1$ and the error term referred to as $u_{i,t}$. The error term is uncorrelated with f_t. The key assumptions of the static factor model, as discussed in Nard et al. (2021), are as follows: the factor loadings β_i are time-invariant, and the conditional covariance matrix of the factors f_t and the errors u_t are time-invariant. The factor model can be expressed in matrix notation as:

$$e_t = Bf_t + u_t$$

Where B is a $p \times K$ matrix that contains the β_i vectors. To obtain the covariance matrix of the forecast errors, Σ_e, the following equation is used:

$$\Sigma_e = B' \mathrm{cov}(f_t) B + \Sigma_u \tag{20}$$

Where Σ_u is the covariance matrix of u_t.
Since the factors are unknown, an estimation method developed by Fan, Liao, et al. (2013) is followed. The data is utilized through the sample covariance matrix $\widehat{\Sigma}_{sample}$ of the forecast error e_t. To infer the factors, a singular value decomposition (SVD) is applied to $\widehat{\Sigma}_{sample}$. The resulting covariance matrix formed by the first K principal components is kept. A threshold procedure is then applied to the remaining covariance matrix. This procedure leads to the principal orthogonal complement threshold estimator (POET), which is the final covariance matrix estimate. The theoretical details of the POET estimation method can be found in Fan, Liao, et al. (2013), providing a comprehensive explanation of the procedure.

3.3 Conclusion Performance-improving Techniques

This part explored different methods to improve the estimation of the forecast combination. Firstly, different trimming methods have been discussed from Radchenko et al. (2023). The trimming method aims to reduce the variance of the weights while it will also increase the bias, a bias-variance trade-off.
Secondly, various covariance matrix estimators have been explored. The sample covariance estimator is often used. This method is unbiased and is the maximum likelihood estimator under normality. It may not be the optimal estimator in high-dimensional settings. The sample covariance estimate could be numerically ill-conditioned. Inverting this matrix can increase estimation errors.
An alternative covariance estimation method is the linear shrinkage method. This method combines the sample covariance matrix with a structured target matrix, such as the identity matrix. The aim is to reduce the impact of noise and increase stability. This could lead to a more efficient covariance matrix estimator in the high-dimensional setting.
The last alternative covariance estimation method that is discussed in this section is the factor model. This model is a widely used finance to estimate the covariance matrix in a high-dimension setting. This makes it a potential proper method for determining the covariance matrix in the forecast combination setting.

4 Trimming Method in Forecast Combination

In Section 3, different methods were explored to enhance the estimation of a forecast combination. This section aims to evaluate the effectiveness of the first method, the trimming method in a forecast combination setting that allows for negative weights. The functionality of the trimming method is evaluated through a simulation study and an empirical application similar to the one conducted by Radchenko et al. (2023). The aim is to replicate this study.

The first section will provide the framework and present the results observed in the simulation study. In the subsequent section, the trimming methods will be applied in the empirical application using inflation forecasts (HICP), real GDP growth forecasts (RGDP), and the unemployment rate forecasts (UNEM) from the European Central Bank (ECB) Survey of Professional Forecasters (SPF) . Differences between the obtained results and those of Radchenko et al. (2023) will be carefully examined and discussed.

4.1 Monte Carlo Simulation

The study of Radchenko et al. (2023) demonstrates the trimming effect in estimating negative weight for forecast combination using a Monte Carlo simulation with two forecasts. This simulation is based on Smith and Wallis (2009) with the necessary modifications to ensure the optimal weight is negative.

4.1.1 The Set-up

The development of the simulation is shortly discussed in Radchenko et al. (2023). In this part, their steps are retaken and will be reviewed in more detail. The simulation in the paper of Smith and Wallis (2009) is an AR(2) process, generating the observations z_t. Two cases are studied in depth. To ensure that negative weights become optimal, the following inequality needs to hold: $\rho > \sigma_2/\sigma_1$. Where ρ denotes the forecast error correlation and σ_i is the forecast error variance for forecast i. This equation can be reformulated as $\text{cov}(e_1, e_2) > \sigma_2^2$, with e_i the forecast error for forecast i. For an AR(2) process this inequality holds for different ranges of ϕ_1 and ϕ_2, depending on their relationship. However, using an AR(1) model allows for a more comprehensive selection of the range for the ϕ_1 parameter and the two forecaster methods. This is the first adaption made in the simulation.

The negative weight simulation still needs two forecasts. The forecast functions discussed in the two cases in Smith and Wallis (2009) can be considered as candidates, provided they satisfy the $\rho > \sigma_2/\sigma_1$ inequality. The forecasts discussed in this paper are transformed into an AR(1) form. In the first case, the two forecasts are derived as $y_{1t} = z_{t-1}$ and $y_{2t} = (2\phi_1 - 1)z_{t-1}$. Selected to ensure that both forecasts have the same forecast error variance. In the second case, the two forecasts are based on $y_{1t} = \phi_1 z_{t-1}$ and $y_{2t} = \phi_1^2 z_{t-2}$. Both forecasts are unbiased in this case. After assessing different combinations of the forecasts of both cases, the choice is made to use $y_{1t} = z_{t-1}$ from case 1 and $y_{2t} = \phi_1^2 z_{t-2}$ from case 2, as they satisfy the requirement for negative optimal weights for all $\phi_1 < 0$.

4.1.2 Theoretical Analysis

Now it is time to look into the simulation in more detail. As discussed, in this study, a time series is generated based on an AR(1) process:

$$z_t = \phi_1 z_{t-1} + \eta_t \qquad t = 1, \ldots, T+1$$

The error η_t are independent and identically distributed from a standard normal distribution. The parameter ϕ_1 is chosen to satisfy $|\phi_1| < 1$. For this given parameter the variance of the process can be determined as:

$$\sigma_z^2 = var(z_t) = \frac{1}{1+\phi_1^2}$$

Additionally, the first two autocorrelation coefficients are given by:

$$\rho_1 = \text{corr}(z_t, z_{t-1}) = \phi_1 \qquad\qquad \rho_2 = \text{corr}(z_t, z_{t-2}) = \phi_1^2$$

In this simulation study, the aim is to forecast z_{T+1}, using the two forecast methods discussed before:

$$y_1 = z_T \qquad\qquad y_2 = \rho_2 z_{T-1}$$

The properties of these forecast combinations $y_c = wy_1 + (1-w)y_2$, with weight w, are investigated for different values of ϕ_1. The weights are estimated using the first T observations, with $T = 30$ in this simulation. The errors are defined as $e_{1t} = z_t - z_{t-1}$ and $e_{2t} = z_t - \rho_2 z_{t-2}$. This leads to forecast errors:

$$\begin{aligned} e_1 &= e_{1,T+1} = z_{T+1} - y_1 & \qquad e_2 &= e_{2,T+1} = z_{T+1} - y_2 \\ e_1 &= z_{T+1} - z_T & e_2 &= z_{T+1} - \rho_2 z_{T-1} \end{aligned}$$

The variances of these forecast errors are:

$$\sigma_1^2 = \text{var}(e_1) = 2(1-\rho_1)\sigma_z^2 \qquad\qquad \sigma_2^2 = \text{var}(e_2) = (1-\rho_2^2)\sigma_z^2$$

The covariance between the forecast errors is:

$$\text{cov}(e_1, e_2) = \sigma_z^2(1-\rho_2)(1-\rho_1+\rho_2)$$

The variance and covariance expressions are derived in detail in Appendix A. Using the covariance, the correlation is computed by $\rho = \text{cov}(e_1, e_2)/\sigma_1\sigma_2$. In Subsection 2.4 , it was established that the optimal weight is negative when $\rho > \frac{\sigma_2}{\sigma_1}$. As discussed at the beginning of this part, this holds for $\phi_1 < 0$.

Provided with the true covariance matrix of the forecast errors, Σ, the theoretical optimal weights for the forecast combination $y_c = w'\mathbf{y}$ are determined by the optimal weight equation, seen in Equation 4. Following Radchenko et al. (2023), in the simulation the covariance matrix will be estimated by the sample covariance matrix, $\widehat{\Sigma}$, by method (16). The error means are defines as: $\bar{e_1} = \frac{1}{T-2}\sum_{t=2}^{T-1} e_{1,t+1}$ and $\bar{e_2} = \frac{1}{T-2}\sum_{t=2}^{T-1} e_{2,t+1}$. With these error means, the sample covariance matrix is rewritten and calculated as:

$$\begin{aligned} \widehat{\Sigma} &= \begin{pmatrix} \widehat{\sigma}_1^2 & \widehat{\rho}\widehat{\sigma}_1\widehat{\sigma}_2 \\ \widehat{\rho}\widehat{\sigma}_1\widehat{\sigma}_2 & \widehat{\sigma}_2^2 \end{pmatrix} \\ &= \frac{1}{T-3} \times \sum_{t=2}^{T-1} \begin{pmatrix} (e_{1,t+1} - \bar{e_1})^2 & (e_{1,t+1} - \bar{e_1})(e_{2,t+1} - \bar{e_2}) \\ (e_{1,t+1} - \bar{e_1})(e_{2,t+1} - \bar{e_2}) & (e_{2,t+1} - \bar{e_2})^2 \end{pmatrix} \end{aligned}$$

With the sample covariance matrix, the theoretically optimal weight w^* is estimated using method (1), by $\widehat{w}$:

$$\widehat{w} = \frac{\widehat{\sigma}_2^2 - \widehat{\rho}\widehat{\sigma}_1\widehat{\sigma}_2}{\widehat{\sigma}_1^2 + \widehat{\sigma}_2^2 - 2\widehat{\rho}\widehat{\sigma}_1\widehat{\sigma}_2}$$

Since there are two forecasts available, the trimming methods discussed in Subsection 3.1 will result in the same weight. Therefore, the weights are determined by:

$$w^{\text{TR}}(c) = \max(\widehat{w}, -c) = \begin{cases} \widehat{w}, & \widehat{w} > c \\ c, & \widehat{w} \leq c \end{cases}$$

4.1.3 Performance

To determine the impact of different threshold values, c, a simulation is performed with a range of values between -1 and 0. This simulation is repeated 1 million times, yielding vectors $\widehat{w}$ and w^{TR} and the forecast error $e_c = z_{T+1} - y_c$. Afterward, the Mean Squared Forecasting Error (MSFE) is calculated as $\mathrm{MSFE} = \frac{1}{n}\sum_{i=1}^{n}(e_c)^2$.
Since the Data Generating Process (DGP) and the forecast methods are known, it becomes possible to calculate the theoretically optimal weight w^*. These weights are obtained by substituting the true covariance matrix, Σ, into Equation 4. Where Σ is derived from the theoretical variances and the covariance of the forecast errors. For the two forecasts discussed in this simulation with parameter $\phi_1 = -0.5$, the theoretical optimal weight is calculated as $w^* = -0.2857$ and complementary $1 - w^* = 1.2857$. The smallest MSFE is observed for $c = -0.26$, which is close to the theoretical optimal weight of $w^* = -0.29$. The bias resulting from the trimming will therefore be small. This result corroborates the findings from Radchenko et al. (2023). The variance of the estimated optimal weights without trimming is 0.0134. The trimming method reduces the variance of the trimmed weights to 0.0053. Reducing the trimming value to zero results in a variance of 0.000158. These results differ slightly from Radchenko et al. (2023), where the reported variance of the optimal untrimmed weights is 0.0125 and with trimming reduced to 0.0049. Trimming at zero reduces the variance in their study to 0.0001. The conclusion is still the same. The negative trimming method will cause some bias, however, the variance of the weights is notably reduced. This offsets the bias effect and will generate better performing forecasts.
Figure 1 illustrates the MSFE plot for $\phi = -0.5$. The right plot shows the MSFE value for different threshold values. The red vertical line indicates the estimated optimal trimming threshold of -0.26. The left plot features the density function of the estimated weights before trimming by the dashed light green line. The weights are currently estimated, but it is also possible to use fixed weights. A fixed means that the weight is set to a specific value without estimation. The MSFE is also determined for a range of fixed weights, this is represented by the solid blue line in the plot. If the weight is fixed at 0, the MSFE value using only the second forecast is calculated, indicated by σ_2 in the plot. Similarly, a fixed weight of 1 represents the MSFE value using only the first forecast, indicated by σ_1 in the plot. The MSFE of the theoretical optimal weight of -0.29 is represented by point F. These results closely resemble the results reported by Radchenko et al. (2023).
The MSFE function for different ϕ_1 values between -0.9 and 0 is seen in Figure 2. The optimal weight is negative if $\phi_1 < 1$, indicating the possibility for a negative optimal trimming value. In both plots in Figure 2 the optimal trimming value is indicated by the light blue dot on the MSFE curves. The figure closely resembles the figure in Radchenko et al. (2023).

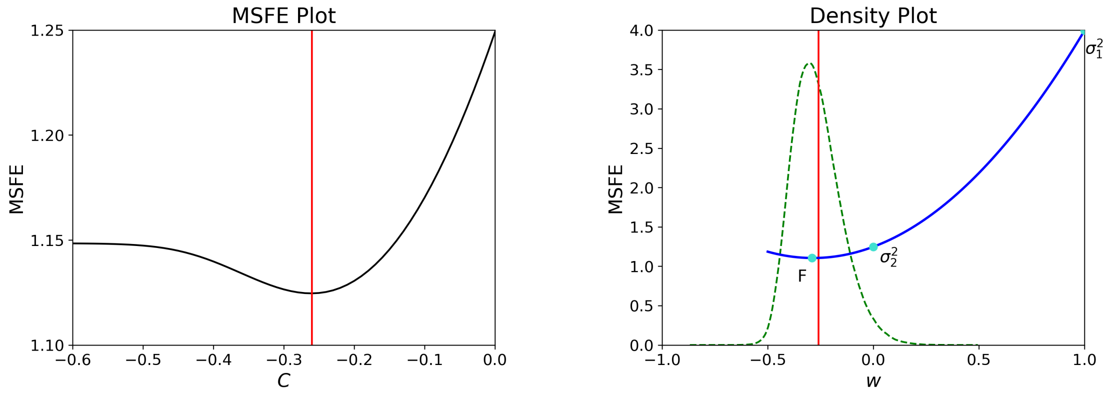

Figure 1: MSFE plot for the simulation with autoregressive parameter $\phi_1 = -0.5$. The left plot displays the MSFE values for different trimming thresholds of c between -0.6 and 0. The red vertical line indicates the threshold value that minimized the MSFE of $c = -0.26$. The right plot presents the distribution of the untrimmed weights, $\hat{w}$, depicted by the dashed green line. The blue line represents the MSFE value given a fixed weight, with point F representing the theoretically optimal weight $w^* = -0.29$. The two σ points denote the original single forecasts.

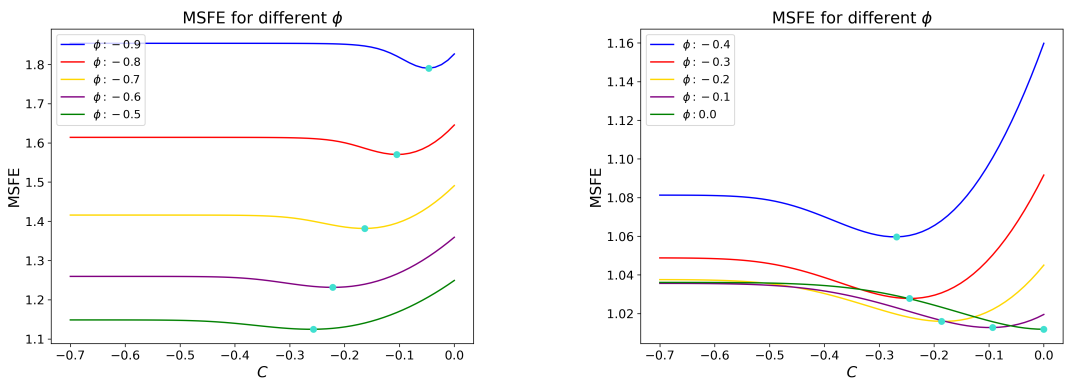

Figure 2: The MSFE function of trimming thresholds c for values for autoregressive parameter ϕ_1 between -0.9 and 0. The left plot contains the curves for ϕ_1 between -0.9 and -0.5 with steps of 0.1. The right plot shows the curves for ϕ_1 between -0.5 and 0.The light blue point indicates the optimal trimming threshold.

4.2 Empirical application

Allowing for negative weights alongside the use of different trimming methods are applied to quarterly forecasts of inflation (HICP), real GDP growth (RGDP), and unemployment rate (UNEM) using the data of the European Central Bank (ECB) Survey of Professional Forecasters (SPF). A similar approach as seen in Matsypura et al. (2018) is followed. Likewise, the one- and two-year-ahead forecast horizons from Q4 1999 to Q2 2018 will be used. In this data set, there are 100 forecasters. The forecasters that participated for at least 24 quarters, 6 years, are selected to avoid outliers and ensure robustness. The forecasts are evaluated based on the ECB macroeconomic indices, following Radchenko et al. (2023). The data is publicly available.

4.2.1 Preliminary Analysis

In Subsection 2.4, it was observed that negative weights are more prevalent when there is a high correlation among different forecasters and when all the forecasts consistently overestimate or underestimate the true value. In the study conducted by Radchenko et al. (2023), both patterns became visible for variables HICP, RGDP, and UNEM, supporting the expectation of the presence of negative weights. This belief is validated by a review of the distribution of estimated optimal weights. As outlined in Subsection 2.2, the optimal weights are determined by Equation 4. The methodology presented by Matsypura et al. (2018) is used to estimate the covariance matrix Σ by an adapted version of the sample covariance method (16). For each pair of forecasters, the time instances of overlap are determined. Based on this sample of corresponding forecasts the covariance is determined. Let $\mathcal{T}_i \subseteq \{1, 2, \ldots, T\}$ represent the set of time instances that forecaster i has a forecast available. The covariance matrix is then estimated as:

$$\widehat{\sigma}_{ij} = \begin{cases} \dfrac{1}{|\mathcal{T}_i|} \displaystyle\sum_{t \in \mathcal{T}_i} [y_{it} - E(y_{it})]^2, & \text{if } i = j \\ \dfrac{1}{|\mathcal{T}_i \cap \mathcal{T}_j|} \displaystyle\sum_{t \in \mathcal{T}_i \cap \mathcal{T}_j} [y_{it} - E(y_{it})][y_{jt} - E(y_{jt})], & \text{if } i \neq j \end{cases}$$

In situations where there is no overlap in time instance, the covariance value between forecasters becomes zero. The covariance matrix needs to be invertible, and therefore positive definite. To make sure this is the case, the nearPD function is used in R. This function computes the positive definite matrix closest to a given matrix. It is important to note that applying a different function that computes the nearest positive definite matrix will result in a matrix, leading to different weight estimations. The use of the nearPD function in R is essential to obtain similar results.

In Table 1, the quantiles of all optimal weights for all forecasters throughout the entire forecasting period are presented. In this table, a significant portion of the distribution of optimal weights is negative across all types of forecasts. This suggests the potential benefits of using a negative trimming threshold, rather than zero. The difference between the obtained results and those of Radchenko et al. (2023) is minimal. In a few cases, such as the optimal weights for the 1-year HICP forecasters, the values at the 0.0 quantiles differ only 0.0001. For all other cases, the results are exactly the same.

Table 1: The quantiles of all estimated optimal weights of all forecasters for the entire forecasting period.

	1-year			2-year		
Quantile	HICP	RGDP	UNEM	HICP	RGDP	UNEM
0.0	−9.3224	−9.0727	−14.4107	−10.1254	−30.0941	−3.2778
0.1	−0.7105	−0.6153	−0.6554	−0.6201	−0.9821	−0.6978
0.2	−0.3827	−0.2678	−0.3571	−0.3437	−0.5043	−0.3319
0.3	−0.2142	−0.1335	−0.1996	−0.2317	−0.2952	−0.1744
0.4	−0.1108	−0.0326	−0.0970	−0.0889	−0.0983	−0.0632
0.5	−0.0039	0.0355	0.0149	0.0183	0.0683	0.0131
0.6	0.1256	0.1094	0.1163	0.1112	0.2223	0.1402
0.7	0.2563	0.1913	0.2343	0.2401	0.4001	0.2750
0.8	0.4365	0.3096	0.4415	0.4227	0.6264	0.5092
0.9	0.8556	0.6402	0.7600	0.7277	1.0094	0.7625
1.0	11.0676	9.1278	11.2623	7.8713	25.4852	2.7262

4.2.2 Fixed Trimming Threshold

As mentioned in the previously, Table 1 presents results indicating the presence of negative weights. In the Monte Carlo simulation study, the use of trimmed weights resulted in improved performance. Since two forecasts were present, the trimming method was of no influence. However, the empirical study involves more forecasters compared to the simulation study. Consequently, the trimming method now has an impact on the resulting optimal weight vector due to the increased number of forecasters. To assess the effectiveness of trimming, five different trimming methods will be evaluated. These methods are discussed in Subsection 3.1. The range of the trimming threshold will be from $-\infty$ to 0, where $-\infty$ results in no trimming. These five trimming methods will be evaluated based on the mean squared forecasting error (MSFE). A fixed value for the trimming parameter c will be chosen for all 16 testing periods, and the MSFE will be averaged across these periods.

It is important to note that fixing the trimming value for all testing periods may not be optimal, as different testing periods might require different trimming thresholds. However, this initial evaluation provides an indication of whether trimming negative weights could contribute to an improvement in forecast combination in the empirical setting. Subsequently, a data-driven method with flexible trimming thresholds will be assessed to further evaluate the impact of trimming on forecast combination performance.

In order to assess whether the trimming method outperforms the standard equal weights method, the ratio of the MSFE using trimmed weights and the MSFE of equal weights is examined, referred to as the relative MSFE. If this ratio is less than one, it indicates that the trimming method outperforms the equal weights method. To test for statistical significance, the Diebold-Mariano (DM) test is used, developed in Diebold and Mariano (1995).

Table 10 in Appendix B the results for trimming methods 1, 2, and 3 are presented, including the p-values of the DM test in brackets. Since the same data from the ECB SPF is used and trimming methods 1, 2, and 3 have no stochastic component, the results match exactly with those presented in Radchenko et al. (2023). Table 2 shows the results for trimming methods 4 and 5. In this case, there are some differences compared to Radchenko et al. (2023). Since the initial weights are randomly chosen for trimming methods 4 and 5, this is also expected. Despite these differences, the overall pattern regarding the performance of the trimming method remains the same.

Equal weights still outperform negative weights when no trimming is applied, $c = -\infty$, except for the one-year ahead estimate of RGDP. The estimated optimal weights without trimming do not outperform equal weights probably due to the large estimation errors of the weights, as is shown in Subsection 2.3. However, the large p-values suggest that the differences are not statistically significant. Trimming at $c = 0$ often leads to improved performance compared to equal weights. Using trimming methods 4 and 5 results in an improvement with confidence, except for the two-year ahead estimates of HICP and UNEM. Although the relative MSFE values differ from Radchenko et al. (2023) for trimming methods 4 and 5, it is worth noting that trimming close to zero often results in a lower MSFE compared to trimming at zero. This is particularly notable in the case of the UNEM forecast for the one-year horizon. This is an indication that trimming at a negative weight could potentially improve the forecast combination.

Table 2: The relative MSFE values of the forecast combination using trimming methods 4 and 5 with various trimming thresholds.

	1-year			2-year		
TR4	HICP	RGDP	UNEM	HICP	RGDP	UNEM
c: $-\infty$	8.959(0.139)	0.874(0.470)	1.065(0.870)	3.814(0.105)	5.339(0.333)	4.625(0.191)
-5.0	3.445(0.058)	1.050(0.886)	3.651(0.098)	3.470(0.047)	2.159(0.149)	19.801(0.158)
-4.5	9.625(0.010)	0.953(0.881)	2.819(0.085)	3.415(0.021)	4.071(0.175)	14.116(0.046)
-4.0	5.849(0.134)	0.971(0.921)	2.871(0.022)	3.898(0.059)	0.888(0.783)	7.955(0.022)
-3.5	3.116(0.022)	1.501(0.234)	3.601(0.058)	3.719(0.027)	2.058(0.285)	11.203(0.085)
-3.0	2.892(0.094)	1.303(0.322)	2.690(0.105)	2.699(0.021)	1.084(0.842)	12.498(0.084)
-2.5	4.584(0.036)	1.536(0.207)	1.848(0.102)	2.546(0.039)	1.296(0.485)	5.342(0.110)
-2.0	2.432(0.031)	0.779(0.276)	1.309(0.462)	2.877(0.009)	1.103(0.776)	4.563(0.082)
-1.5	3.818(0.110)	0.906(0.593)	1.855(0.195)	2.258(0.027)	1.299(0.447)	3.416(0.064)
-1.0	2.283(0.060)	0.729(0.096)	0.746(0.469)	1.782(0.067)	1.084(0.773)	1.916(0.211)
-0.5	1.891(0.090)	0.648(0.007)	0.532(0.046)	1.668(0.060)	0.712(0.133)	0.908(0.755)
0.0	0.865(0.095)	0.893(0.001)	0.695(0.001)	0.902(0.160)	0.810(0.001)	0.976(0.856)
TR5						
c: $-\infty$	8.959(0.139)	0.874(0.470)	1.065(0.870)	3.814(0.105)	5.339(0.333)	4.625(0.191)
-5.0	1.695(0.171)	0.887(0.223)	0.722(0.226)	1.238(0.431)	0.567(0.003)	1.858(0.111)
-4.5	1.688(0.110)	0.741(0.043)	0.496(0.002)	1.399(0.149)	0.548(0.001)	1.645(0.197)
-4.0	1.529(0.166)	0.772(0.009)	0.406(0.000)	1.321(0.191)	0.545(0.000)	1.488(0.344)
-3.5	1.477(0.164)	0.772(0.010)	0.373(0.000)	1.176(0.376)	0.539(0.000)	1.316(0.499)
-3.0	1.445(0.170)	0.772(0.026)	0.323(0.000)	1.107(0.481)	0.536(0.000)	1.103(0.789)
-2.5	1.236(0.354)	0.778(0.017)	0.256(0.000)	1.037(0.725)	0.538(0.000)	0.933(0.829)
-2.0	1.134(0.492)	0.787(0.023)	0.236(0.000)	1.024(0.783)	0.566(0.000)	0.748(0.346)
-1.5	1.028(0.849)	0.831(0.021)	0.199(0.000)	0.993(0.931)	0.603(0.000)	0.577(0.077)
-1.0	0.937(0.615)	0.872(0.029)	0.199(0.000)	0.970(0.729)	0.654(0.000)	0.474(0.010)
-0.5	0.896(0.320)	0.889(0.005)	0.303(0.000)	0.960(0.616)	0.740(0.000)	0.530(0.003)
0.0	0.864(0.094)	0.893(0.001)	0.695(0.001)	0.902(0.160)	0.810(0.001)	0.976(0.856)

Note: In order to assess whether the trimming method outperforms the standard equal weights method, the ratio of the mean squared forecast error (MSFE) using trimmed weights and the MSFE of equal weights is examined, referred to as the relative MSFE. This table presents the results for these different two different forecast combination methods for different trimming thresholds c, indicated on the left side. The MSFE is averaged across the testing period. If this ratio is less than one, it indicates that the trimming method outperforms the equal weights method. To test for statistical significance, the two-sided DM test is used. The p-values are shown between the brackets.

Overall, trimming method 5 performs the best in this empirical setting, resulting in a forecast combination that performs 80% better than equal weights for the UNEM one-year ahead forecast and a 50% improvement for the UNEM two-year ahead forecast. This leads to the conclusion that the trimming at zero method could improve the forecast combination compared to equal weights, and a negative trimming value might further improve the forecast combination. This finding strongly suggests that negative weights can contribute to forecast combinations. This will be further explored using a more realistic setting involving a data-driven dynamic threshold instead.

4.2.3 Data-driven Trimming Threshold

The previous subsection looked into a fixed threshold setting where for all 16 testing periods the same trimming value was used. However, in practice, the optimal threshold may vary across different periods. Based on the Equation (14) of pseudo-out-of-sample AMSFE, explained in Subsection 3.1.3, a different threshold value will be determined for every testing period. The threshold will be selected between c_{min} and 0 with a step size of 0.1 using Equation (15). In case of different thresholds resulting in the same AMSFE, the largest will be picked. Two cases are evaluated, $c_{min} = -5$ and $c_{min} = -2$.

Table 3 presents the results of the data-driven threshold method. As before, the relative MSFE is shown, indicating the MSFE of the threshold method divided by the equal weights method. The significance is assessed using the DM test. For the one- and two-year ahead forecasts of HICP, no significant differences are observed between the methods. However, for the one- and two-year ahead forecasts of RGDP, trimming methods 4 and 5 demonstrate a significant improvement compared to the equal weights method. Similarly, a significant improvement is observed for the one-year ahead forecast of UNEM across all trimming methods. For the two-year ahead forecast of UNEM, the results are less conclusive in terms of significance.

Table 3: The relative MSFE values of the forecasting combination using the data-driven threshold method based on the pseudo out-of-sample evaluation.

	1-year			2-year		
$c_{min} = -2$	HICP	RGDP	UNEM	HICP	RGDP	UNEM
TR1	1.243(0.095)	0.896(0.045)	0.813(0.009)	1.037(0.557)	0.880(0.129)	0.832(0.153)
TR2	1.166(0.177)	0.931(0.204)	0.812(0.013)	1.040(0.486)	0.922(0.333)	0.881(0.499)
TR3	1.161(0.123)	0.912(0.063)	0.761(0.007)	1.072(0.264)	0.901(0.280)	0.784(0.094)
TR4	0.865(0.095)	0.893(0.001)	0.702(0.010)	0.873(0.124)	0.810(0.001)	0.957(0.719)
TR5	0.869(0.111)	0.884(0.002)	0.462(0.000)	0.907(0.209)	0.781(0.001)	0.866(0.239)
$c_{min} = -5$						
TR1	1.243(0.095)	0.900(0.055)	0.823(0.019)	1.037(0.557)	0.876(0.121)	0.832(0.153)
TR2	1.166(0.177)	0.931(0.204)	0.812(0.013)	1.040(0.486)	0.922(0.333)	0.881(0.499)
TR3	1.161(0.123)	0.912(0.063)	0.761(0.007)	1.072(0.264)	0.901(0.280)	0.784(0.094)
TR4	0.865(0.094)	0.893(0.001)	0.702(0.010)	0.873(0.124)	0.810(0.001)	0.957(0.719)
TR5	0.869(0.111)	0.884(0.001)	0.458(0.000)	0.907(0.209)	0.781(0.001)	0.866(0.240)

Note: In order to assess whether the data-driven threshold method outperforms the standard equal weights method, the ratio of the mean squared forecast error (MSFE) using trimmed weights and the MSFE of equal weights is examined, referred to as the relative MSFE. This table presents the results for these different two different forecast combination methods. The MSFE is averaged across the testing period. If this ratio is less than one, it indicates that the data-driven threshold method outperforms the equal weights method. To test for statistical significance, the two-sided DM test is used. The p-values are shown between the brackets.

The relative values obtained in the thesis are generally similar to those reported in Radchenko et al. (2023). However, several differences in the reported p-values have been identified. Radchenko et al. (2023) indicates significance using stars without providing precise values, so the exact difference is not possible to determine. For trimming methods 1 and 2 for the two-year ahead forecast of RGDP they find a 1% significant difference between the data-driven MSFE and the equal weights MSFE. This thesis determines a p-value of another order, indicating the opposite of a significant result. After inspecting their code, several errors in their evaluation of p-values for the data-driven threshold method were discovered. However, the p-value calculations for the fixed threshold method are accurate and match the p-values determined in the thesis.

In Appendix B, additional results are presented for the data-driven threshold method. These include Table 11 showing the optimal trimming thresholds for each testing period across all trimming methods for the one-year ahead forecast of UNEM and the one-year ahead forecast of HICP. These trimming thresholds are determined by a pseudo out-of-sample evaluation. In this table, the c_{min} is set at -2. It is noteworthy that for the HICP case, the optimal threshold is often set at

0. This aligns with the expectation since the fixed threshold already suggests that a threshold of 0 performs better than smaller thresholds in many instances. For the UNEM case, more frequent optimal thresholds below zero are observed. Again, this is in line with the expectation, as in the fixed threshold setting, a threshold of -0.5 often outperforms a threshold of 0 indicating negative weights could be beneficial.

In Figure 6 in Appendix B a histogram is presented of the weights obtained from trimming method 3 of the one-year ahead forecast of HICP and the histogram of the one-year ahead forecast of UNEM across the 16 testing periods. The impact of the dynamic threshold is clearly noted. In the histogram negative weights are more frequent in the UNEM histogram compared to the HICP histogram.

4.3 Conclusion of the Performance of Trimming

This thesis confirms the results reported by Radchenko et al. (2023), with the only difference being in trimming methods 4 and 5 due to their stochastic components. Additionally, a coding mistake is identified in the determination of the p-values for the data-driven threshold method, a correct version of the R code is publicly available as well.[3] As concluded by Radchenko et al. (2023), the trimming method outperforms the use of optimal weights due to large estimation errors. Trimming reduces the variance of the weights while resulting in some additional bias. However, the effect of the variance reduction outweighs the bias effect. This leads to the idea to find other ways to reduce the estimation errors. One potential method is to improve the estimation of the covariance matrix since most estimation errors in the current method are likely due to the sample covariance estimation. In high-dimensional settings like this empirical application, the sample covariance method is known to have high estimation errors. Fortunately, there are several methods available in the literature to improve covariance estimation and potentially reduce these errors. In the next part of the thesis, these methods will be evaluated for their potential for improving forecast combination performance.

5 Covariance Estimation Methods in Forecast Combination

In the empirical application of Radchenko et al. (2023), the trimming method is employed to forecast inflation (HICP), real GDP growth (RGDP), and the unemployment rate (UNEM) using data from the European Central Bank (ECB) Survey of Professional Forecasters (SPF). A crucial step in this process is the estimation of the covariance matrix. In this empirical setting, this is challenging due to the large number of forecasters and the relatively small dataset of quarterly information available. This leads to a large covariance matrix needed to be estimated. In Radchenko et al. (2023) the sample covariance estimation method is used. However, this method is known for its problems in a high-dimensional setting.

In order to determine the performance of the different covariance estimation methods, a simulation study will be conducted, inspired by the setup in Radchenko et al. (2023) with a larger number of forecasts. This study aims to evaluate the performance of these alternative estimation methods in a controlled setting.

Finally, these alternative methods will be applied to the empirical application of Radchenko et al. (2023). The performance of these alternative covariance estimation methods will be compared to the equal weights method. Additionally, the performance will be compared to the data-driven

[3]See `https://github.com/YannickvanEtten/bachelor_thesis` for the R code for the data-driven threshold with correct p-values.

threshold method derived in Radchenko et al. (2023) as well. This comparison will provide insights into the effectiveness of different covariance matrix estimation methods and their potential improvements. This can be seen as an extension of Radchenko et al. (2023).

5.1 High-dimensional Simulation

In a high-dimensional situation, such as the empirical application, the use of the sample covariance method could lead to problems. To examine possible improvements, alternative estimation methods are considered, namely the linear shrinkage method and the factor model discussed in Subsection 3.2. In this subsection, these alternative methods are evaluated in a high-dimensional simulation.

5.1.1 Simulation Setting

The effectiveness of these estimation methods and their suitability for the negative weight estimation is assessed by a Monte Carlo simulation inspired by Radchenko et al. (2023). Consequently, there will be some overlapping elements in the description compared to Subsection 4.1.2. The objective is to evaluate different covariance estimation methods in a high-dimensional setting, aiming to create a similar scenario as observed in the empirical application.
Following the replicated simulation study, the time series data is generated using an AR(1) process:

$$z_t = \phi_1 z_{t-1} + \eta_t \qquad t = 1, \dots, T+1$$

Where the autoregressive parameter, ϕ_1 is chosen such that $|\phi_1| < 1$ to ensure the stability of the process. The auto-correlation coefficient function ρ_i is defined as the correlation between observations z_t and z_{t-i}, given by:

$$\rho_i = \text{corr}(z_t, z_{t-i}) = \phi_1^i$$

In the previous simulation study, only the values of ρ_1 and ρ_2 were provided and used to generate the first two forecasts:

$$y_1 = z_T \qquad\qquad y_2 = \rho_2 z_{T-1}$$

However, in this current simulation study, 23 additional forecasts are included. These additional forecasts, denoted as y_i for $i \in 3, \dots, 25$, are based on the unbiased forecast y_2. Specifically, the formula for generating these additional forecasts is:

$$y_i = \rho_i z_{T-i}$$

This results in a total of 25 forecasts. Again the properties of the forecast combination $y_c = w'y$ are investigated, where $y = (y_1, \dots, y_25)$ and w a vector of 25 weights. Since forecasts 2 to 25 have the same unbiased property, they also have a similar forecast error variance structure. The properties of forecast 1 and forecasts $i \in 2, \dots, 25$ will be explicitly examined. The forecast 1 has an error of $e_{1t} = z_t - z_{t-1}$ and the additional forecasts have an error of $e_{it} = z_t - \rho_i$ for $i \in 2, \dots, 25$. The final forecast error is:

$$e_1 = e_{1,T+1} = z_{T+1} - y_1 \qquad\qquad e_i = e_{i,T+1} = z_{T+1} - y_i$$
$$e_1 = z_{T+1} - z_T \qquad\qquad e_i = z_{T+1} - \rho_i z_{T-i}$$

This leads to the following forecast error variances:

$$\sigma_1^2 = \text{var}(e_1) = 2(1-\rho_1)\sigma_z^2 \qquad\qquad \sigma_i^2 = \text{var}(e_i) = (1-\rho_i^2)\sigma_z^2$$

Since there are two types of forecasts, namely y_1 and the unbiased forecasts y_i for $i \in 2, \ldots, 25$, there are two formulas to determine the true covariance. The first formula calculates covariance between forecast error e_1 and the forecast error e_i of an unbiased forecast, given by:

$$\text{cov}(e_1, e_i) = \sigma_z^2(1 - \rho_1 - \rho_i^2 + \rho_i\rho_{i-1})$$

The second formula calculates the covariance between the error of two unbiased forecasts, where $i, j > 1$ and $i \leq j$, given by:

$$\text{cov}(e_i, e_j) = \sigma_z^2(1 - \rho_i^2 - \rho_j^2 + \rho_i\rho_j\rho_{j-i})$$

All the derivations are written out in detail in Appendix C. In addition, the correlation between two forecast errors, e_i and e_j, can be expressed as $\rho_{ij} = \text{cov}(e_i, e_j)/(\sigma_i\sigma_j)$.

In this simulation, the length of the time series depends on the number of forecasts used, as some forecasts utilize observations further back in time. The aim is to have the same number of available error observations as in Radchenko et al. (2023) to estimate the covariance matrix using forecast errors. In their study, 28 error observations were available with sample size $T = 30$, considering the use of z_{t-2} for the y_2 forecast. In the simulation, z_{t-25} is used, so to ensure 28 error observations for all forecasts, the sample size T is set at $T = 28 + 25 = 53$. However, for all forecasts, only the last 28 forecast errors are considered. The error means are calculates as: $\bar{e_i} = \frac{1}{T-25}\sum_{t=25}^{T-1} e_{1,t+1}$.

All necessary tools are present to determine the sample covariance matrix, $\widehat{\Sigma}_{sample}$, by:

$$\widehat{\Sigma}_{sample} = \begin{pmatrix} \widehat{\sigma}_1^2 & \widehat{\rho}_{12}\widehat{\sigma}_1\widehat{\sigma}_2 & \ldots & \widehat{\rho}_{1,25}\widehat{\sigma}_1\widehat{\sigma}_{25} \\ \widehat{\rho}\widehat{\sigma}_1\widehat{\sigma}_2 & \widehat{\sigma}_2^2 & & \\ \vdots & & \ddots & \vdots \\ \widehat{\rho}_{1,25}\widehat{\sigma}_1\widehat{\sigma}_{25} & \ldots & & \widehat{\sigma}_{25}^2 \end{pmatrix}$$

$$= \frac{1}{T-26} \times \sum_{t=25}^{T-1} \begin{pmatrix} (e_{1,t+1} - \bar{e_1})^2 & (e_{1,t+1} - \bar{e_1})(e_{2,t+1} - \bar{e_2}) & \ldots \\ (e_{1,t+1} - \bar{e_1})(e_{2,t+1} - \bar{e_2}) & (e_{2,t+1} - \bar{e_2})^2 & \\ \vdots & & \ddots \end{pmatrix}$$

This sample covariance estimation method is an adapted version of the one seen in Method (16). Two additional methods for estimating the covariance matrix will be employed in this study. The first method is the shrinkage method, which combines the sample covariance matrix with a diagonal matrix. The shrinkage estimator, denoted as $\widehat{\Sigma}_{LS}$, is calculated by Equation (19). This results in: $\widehat{\Sigma}_{LS} = \delta\widehat{\mu_p}I_{25} + (1-\delta)\widehat{\Sigma}_{sample}$, where $\widehat{\mu}_p$ represents the average of the sample variances of all forecasts. Denoted by the average of the diagonal elements of the sample covariance matrix. The shrinkage intensity, δ, will be determined by examining the mean squared forecast error (MSFE) and L2-norm values across various δ values. The goal is to identify the δ value that yields the lowest overall MSFE and lowest L2-norm compared to the true covariance model.

The second alternative method employed is the factor model estimation of the covariance matrix. In this approach, the number of factors is set equal to the number of forecasts, 25. The factor loadings are determined through the singular value decomposition of the matrix consisting of the estimation errors for all forecasts. Using the factor loadings, the factor variances can be estimated. This leads to an estimation of the factor covariance matrix, denoted as $\widehat{\Sigma}_{factor}$, derived in Equation (20).

The weights vector is as usual determined by Equation 4. The first trimming method (8) from Subsection 3.1 is applied to these estimated weights. This trimming method of the estimated weights $\widehat{w}$ is denoted by:

$$\widehat{w}_i^{\text{TR}} = \alpha_1 \begin{cases} \widehat{w}_i, & \widehat{w}_i > -c \\ c, & \widehat{w}_i \leq c \end{cases}$$

Where α_1 is a scalar factor to make sure that the sum of trimmed weights adds up to 1, $(\widehat{w}_i^{\text{TR}})'\mathbf{1} = 1$. The weight vector is determined for each covariance estimation method, resulting in three different combined forecasts. The forecast combination is determined by: $y_{comb} = \mathbf{y} * \widehat{\mathbf{w}}^{\text{TR}}$. Where $\mathbf{y}$ represents the individual forecasts and $\widehat{w}^{\text{TR}}$ the trimmed estimated weight vector. The simulations will repeat the estimation 100000 times, which is fewer than the previous Monte Carlo simulation with a million repetitions. However, these three estimation methods require a lot more calculation power. Given a value of ϕ_1, the MSFE is determined by the average over all iterations in the two simulation settings.
To ensure that negative optimal weights are present, the forecasts need to be correlated. Since there are 25 forecasts dependent on a factor of ϕ, the phi value will need to be $|\phi| \approx 1$. In this simulation $\phi = -0.95$. So the last observation z_{53} has still a correlation of $-0.95^{25} = -0.278$ with 25 observations before.

5.1.2 Performance

As stated before, two simulations will be performed. The first simulation is done to determine the impact of the shrinkage intensity δ of the MSFE function. In this simulation, the optimal shrinkage intensity for the second simulation is determined.
In the first simulation, the weights will be estimated using the shrinkage method. Given the forecast combinations, the forecast error will be determined for each iteration in order to calculate the MSFE. This procedure is done for the following 8 different shrinkage intensities, $\delta \in 0, 0.1, 0, 2, 0.4, 0.6, 0.8, 0.9, 1$. Leading to the linear shrinkage estimator of:

$$\widehat{\Sigma}_{LS} = \delta \widehat{\mu}_p I_p + (1-\delta)\widehat{\Sigma}_{sample}$$

with $\widehat{\mu}_p = \langle \widehat{\Sigma}_{sample}, I_p \rangle$. In Figure 3 the plot with different δ values can be seen. At first glance, it is noted that the shrinkage value of 1, representing the equal weights estimation method, has for all trimming values c, a relatively high MSFE. The curve of $\delta = 0$, representing the sample covariance use, has a high MSFE value too. For all other values of δ the MSFE value is lower. This indicates an increase in the forecasting performance by combining the sample covariance with the scaled identity matrix, instead of using one or the other. The lines of δ values between 0.1 and 0.4 are close to each other. The green line representing $\delta = 0.2$ outperforms the other lines by a little margin.
To verify that the shrinking method performs best with $\delta = 0.2$, the L2-norm of linear shrinkage matrices $\widehat{\Sigma}_{LS}$ and the true covariance matrix, Σ, is calculated. The formula of the L2-norm between matrices X and Y is given by:

$$||X - Y||_2 = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{ij} - Y_{ij})^2}$$

The results can be seen in Table 4. For shrinkage intensity $\delta = 0.2$, the L2-norm is the smallest. Indicating the closest resemblance to the true covariance matrix. This confirms the choice to set the δ to 0.2. However, as said before, the lines of intensities around 0.1 and 0.3 don't differ much.

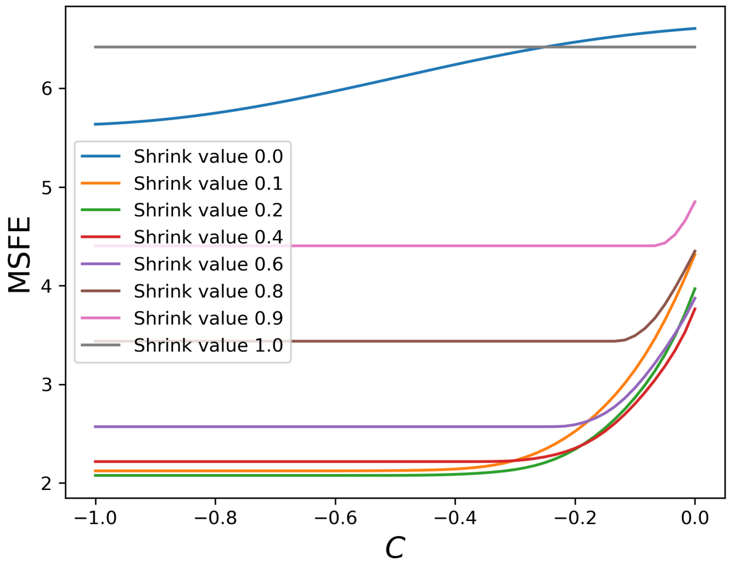

Figure 3: MSFE plot for the simulation of autoregressive parameter $\phi_1 = -0.95$. The plot displays the MSFE functions for different shrinkage parameters δ between 0 and 1 and a trimming threshold c between -1 and 0.

Table 4: The L2-norm between covariance estimate, $\widehat{\Sigma}_{LS}$, and the true covariance matrix, Σ.

Shrinkage intensity δ	L2-norm
0.0	131.580
0.1	126.167
0.2	122.756
0.4	123.351
0.6	136.142
0.8	161.434
0.9	177.285
1.0	194.176

In the first simulation, the best shrinkage intensity has been determined as $\delta = 0.2$. In the second simulation, the MSFE of the three different covariance matrix estimation methods will be calculated for different trimming values. The results are seen in Figure 4.
A clear improvement in the linear shrinkage method and the factor model is noted. Both methods are an improvement to the sample covariance method. The linear shrinkage method outperforms other methods for all trimming thresholds. To get a better understanding of the estimation performance, the L2-norm is calculated for all methods with the true covariance method. The results are seen in Table 5.
The linear shrinkage method has the lowest L2-norm and the best MSFE function performance. Indicating the improvement in the covariance estimation could lead to a decrease in the forecast error.

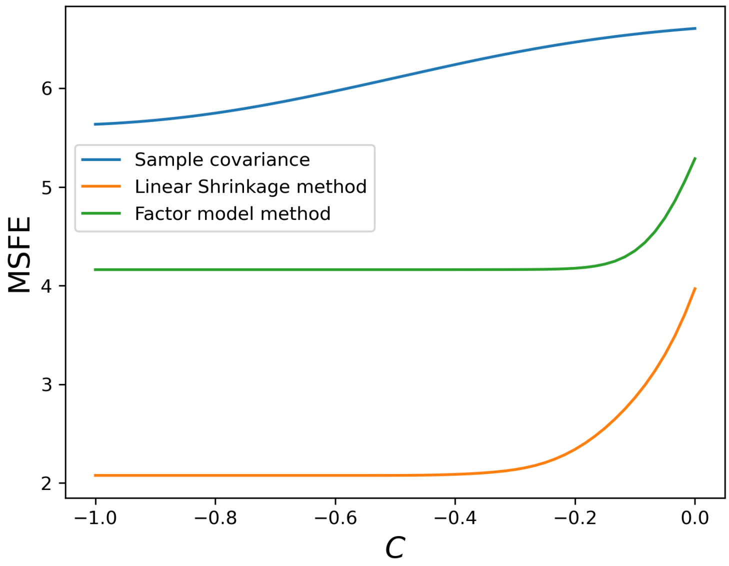

Figure 4: MSFE plot for the simulation of autoregressive parameter $\phi_1 = -0.95$. The plot displays the MSFE functions for different covariance estimation methods: the sample covariance method, the linear shrinkage method and the factor model method. The trimming thresholds c is between -1 and 0. The shrinkage intensity δ is set at 0.2.

Table 5: The L2-norm between different covariance estimates, $\widehat{\Sigma}$, and the true covariance matrix, Σ.

Covariance estimation method	L2-norm
Sample covariance method	131.580
Linear Shrinkage method	122.755
Factor model method	191.504

Furthermore, for the alternative estimation methods, a constant MSFE performance for the factor models when the trimming value c is between -1 and -0.2 and for the linear shrinkage method c between -1 and -0.4 is noted in Figure 4. While a clear upward trend is noticed when c approaches zero, indicating a decrease in the forecast performance. The constant characteristic can be explained by less large negative weights present. If there is no weight available below -0.4, trimming won't have any effect. Since the number of forecasts is increased, the weights could be more distributed among them leading to smaller weight values. The upward trend indicates that the reduction of the variance of the weights does not offset the bias it creates. This means that the trimming method does not have the desired effect, and using the untrimmed weights results in a better forecast. A similar upward trend is seen in the MSFE function of the sample covariance method. This is contrary to the results of the two forecast simulation in Radchenko et al. (2023), indicating that in certain high-dimensional settings, the trimming method might not lead to an improvement.

5.2 Empirical Application with Different Covariance Estimators

In the previous part, two different estimation methods for the covariance matrix were evaluated in a simulation setting. Both these methods led to an improved forecast combination. In this subsection, these methods will be applied to the empirical application of Radchenko et al. (2023). First the linear shrinkage method, thereafter the factor model.

In the empirical application of Radchenko et al. (2023), the dataset initially includes 100 forecasters. After removing forecasters with less than 24 quarters of data, there are 70 forecasters remaining ($p = 70$). The dataset covers quarterly observations from Q4 1999 to Q2 2018, providing 75 data points for HICP, RGDP, and UNEM ($n = 75$). Typically, data partitions are used to estimate the Mean Squared Forecast Error (MSFE), resulting in matrices with dimensions of around 45×59.

In all scenarios considered in Radchenko et al. (2023), $p < n$, suggesting that the estimated covariance matrix should be invertible. However, the ratio of p/n is not negligible, indicating that estimation errors are likely to be amplified. Consequently, using the sample covariance estimation method can result in imprecise estimation results. This is a setting where the curse of dimensionality takes place. This is a perfect place to determine the performance of the alternative methods for covariance matrix estimation.

5.2.1 Empirical Application with the Linear Shrinkage Method

In this part, the linear shrinkage method will be applied to estimate the covariance method of the forecast errors of the SPF, in order to obtain the forecast combination weights. Several experiments are conducted to evaluate the impact of different shrinkage parameters on the mean squared forecasting error (MSFE) in various scenarios. By varying the shrinkage parameter, it is possible to observe how the parameter affects the performance of the forecast combination compared to the equal weights method. The relative MSFE values are plotted against different shrinkage parameter values and analyzed. This allows us to understand the trade-off between bias and variance in the forecast combination. Next, a data-driven approach to determine the optimal shrinkage parameter is introduced. Lastly, the MSFE from the data-driven method for the shrinkage parameter will be evaluated with MSFE from the data-driven method for thresholds seen in the study of Radchenko et al. (2023).

Two different cases are noted when looking at the empirical results of Radchenko et al. (2023) for trimming methods 1,2 and 3 in Appendix B. Cases where the relative MSFE of the trimming weights and equal weights is greater than 1, indicate that the equal weights outperform the trimming method. This holds for the one-year ahead HICP forecast using trimming method 1 with $c = -1$. Another case is where the relative MSFE is smaller than one, indicating that the trimming method outperforms the equal weights method. This holds for the one-year ahead UNEM forecast using trimming method 1 with $c = -1$. For both these cases a plot of the shrinkage method with several shrinkage values will be evaluated. The aim is to see the influence of the shrinkage parameter on the relative MSFE.

As determined in Subsection 3.2.3 when the shrinkage parameter is set to 0, the shrinkage method results in the use of the general sample covariance methods, applied in Radchenko et al. (2023) as well. When the shrinkage parameter is set to 1, the shrinkage method results in the use of the equal weights method. Shrinkage values between 0 and 1, lead to a combination of sample covariance method and a scaled identity matrix.

The plots of the one-year ahead HICP and UNEM forecast using trimming method 1 with $c = -1$ are seen in Figure 5. The plots display the relative MSFE value, determined by the ratio of the MSFE of the forecast combination and the MSFE of the equal weights method. On the right graph the relative MSFE value for HICP for trimming method 1 with trimming value -1. At the shrinkage

parameter of 0, the relative MSFE value of 1.655 is seen, this is the same value in Table 10 since only the sample covariance matrix is used. The value 1.655 indicates that equal weights perform better than the trimming method. At the shrinkage parameter of 1, the relative MSFE value of 1 is seen. This is as expected since this shrinkage parameter results in the use of a scaled identity matrix, equal to the use of equal weights. For shrinkage values between 0.05 and 1 a relative MSFE value below 1 is noted, indicating a better performance of the shrinkage method over the equal weights method. The linear shrinkage method has the best performance at $\delta = 0.14$ with a relative MSFE of 0.898, indicated in the figure by the light blue dot. Noting the sample covariance method has a relative MSFE of 1.655, the linear shrinkage method leads to a large improvement.
On the left the plot of the UNEM is seen, at a shrinkage parameter of 0, again the same value as in Table 10 of 0.555. An improvement is noted in the relative MSFE when changing the parameter. The best performance is at $\delta = 0.04$ with a relative MSFE of 0.263. This is a large improvement when using just a little part of the scaled identity matrix.

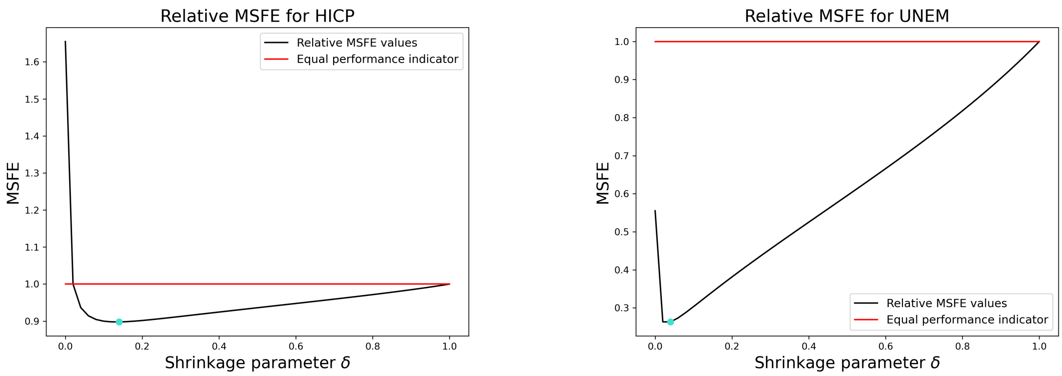

Figure 5: This plot displays the relative MSFE function of the one-year ahead HICP and UNEM forecast using trimming method 1, with threshold $c = -1$. The relative MSFE is determined as the ratio of the MSFE using trimmed weights and the MSFE of equal weights. The linear shrinkage methods is used for the covariance matrix estimation. With values for the shrinkage parameter δ between 0 and 1. The light blue dots represent the optimal relative MSFE value.

As seen in the plots. In both cases, where the sample covariance method with trimming already improves the forecast combination or where the equal weights method still works better, the method of linear shrinkage leads to an increased performance of the forecast combination.
The previous plots indicate linear shrinkage could lead to an increase in the performance of the forecast combinations in many settings. This will be further investigated using different trimming methods. The relative MSFE values are determined by the data-driven linear shrinkage method with regard to the equal weights method. After the data-driven linear shrinkage method to estimate the covariance matrix, trimming methods 1,2, and 3 are implemented. The results are seen in Table 12 in Appendix E. As seen in the table, none of the trimming methods has an impact on the relative MSFE. Therefore, the usage of the trimming methods does not result in an improved forecast combination. It is better to use the estimated weights without trimming, in the table this is seen at $c = -\infty$. This could be due to fewer estimation errors. As determined in Radchenko et al. (2023) trimming reduces the estimation errors of the sample covariance estimation method. With the shrinkage method, the estimation errors should be smaller, as seen in the simulation study. Therefore, trimming should indeed be less impactful. The relative MSFE results of no trimming are seen in Table 6. For the one- and two-year ahead forecast HICP the data-driven trimming

method does not result in a significant improvement. However, for the one- and two-year ahead forecast of the RGDP and the UNEM a significant improvement is seen. Most notably, a 70% improvement for the one-year ahead forecast of UNEM and a 50% improvement for the two-year ahead forecast of UNEM.

The dynamic linear shrinkage method results in an improved forecast combination compared to the equal weights method for 4 of the 6 cases. In the remaining two cases, the dynamic linear shrinkage method delivers similar performance as equal weights. As expected the dynamic linear shrinkage method does not perform worse than the equal weights method. Possibly due to the fact that the equal weights method can be obtained in the dynamic linear shrinkage by setting the shrinkage parameter at 1.

Table 6: The relative MSFE values forecast combination without trimming methods. The weights have been estimated by the linear shrinkage covariance matrix. The shrinkage parameter has been determined by the data-driven method based on the pseudo out-of-sample evaluation.

	1-year			2-year		
	HICP	RGDP	UNEM	HICP	RGDP	UNEM
MSFE	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)

Note: In order to assess whether the shrinkage method outperforms the standard equal weights method, the ratio of the mean squared forecast error (MSFE) using the shrinkage covariance matrix and the MSFE of equal weights is examined, referred to as the relative MSFE. This table presents the results for these different two different forecast combination methods. The MSFE is averaged across the testing period. If this ratio is less than one, it indicates that the shrinkage method outperforms the equal weights method. To test for statistical significance, the two-sided DM test is used. The p-values are shown between the brackets.

The dynamic linear shrinkage leads to a better forecast combination than equal weights in several settings, while it never performs worse. The question remains, does data-driven linear shrinkage result in a better forecast combination than the data-driven trimming methods seen in Radchenko et al. (2023)? This is done by comparing the MSFE values of data-driven linear shrinkage with the MSFE values of the data-driven threshold method. The relative MSFE values of this evaluation are seen in Table 7. The MSFE values of the linear shrinkage method are divided by the data-driven threshold method, so a value below 1 indicates a better performance of the linear shrinkage method. The p-values are as before determined by the DM test and given in brackets.

In Table 7 the difference between the methods for the one- and two-year ahead forecast of HICP and RGDP per trimming method is noted. However, none of the results for the HICP and RGDP has a p-value below 0.05, indicating no clear improvement but also no deterioration. For UNEM a clear significant improvement is noted for most trimming methods, for both the one- and two-year horizon.

Table 7: The relative MSFE values of the forecasting combination using the data-driven linear shrinkage method compared with the data-driven threshold method. For the data-driven shrinkage method, no trimming method is used.

	1-year			2-year		
$c_{min} = -2$	HICP	RGDP	UNEM	HICP	RGDP	UNEM
TR1	0.793(0.071)	1.035(0.416)	0.477(0.005)	0.973(0.577)	0.907(0.351)	0.673(0.055)
TR2	0.845(0.089)	0.996(0.937)	0.478(0.005)	0.970(0.509)	0.866(0.180)	0.635(0.118)
TR3	0.849(0.084)	1.018(0.655)	0.510(0.012)	0.941(0.252)	0.886(0.321)	0.714(0.115)
TR4	1.139(0.085)	1.038(0.189)	0.553(0.002)	1.155(0.080)	0.985(0.588)	0.585(0.015)
TR5	1.130(0.113)	1.049(0.079)	0.841(0.271)	1.112(0.118)	1.022(0.388)	0.646(0.031)

Note: In order to assess whether the data-driven linear shrinkage method outperforms the data-driven trimming threshold method, the ratio of the mean squared forecast error (MSFE) using shrinkage covariance estimation method and the MSFE of the data-driven threshold method is examined, referred to as the relative MSFE. This table presents the results for these different two different forecast combination methods. The MSFE is averaged across the testing period. If this ratio is less than one, it indicates that the data-driven linear shrinkage method outperforms the data-driven trimming threshold method. To test for statistical significance, the two-sided DM test is used. The p-values are shown between the brackets.

5.2.2 Empirical Application with the Factor Method

The factor model is applied to the empirical application of Radchenko et al. (2023) as well. The relative MSFE values are determined by the factor model method with regard to the equal weights method. The optimal weights are estimated using the covariance matrix obtained by the POET estimation method. This method has been developed in Fan, Liao, et al. (2013) and is a covariance matrix estimation based on the factor model. Trimming methods 1,2 and 3 are used on the estimated optimal weights. The results are seen in Table 13 in Appendix E. As seen in the table, none of the trimming thresholds has a positive impact on the relative MSFE. The usage of the trimming methods does not result in an improved forecast combination. It is better to use the estimated weights without trimming, in the table seen at $c = -\infty$. This could be due to fewer estimation errors. As determined by Radchenko et al. (2023) trimming reduces the estimation errors of the sample covariance estimation method. With the factor model method estimation errors should be less present, as also seen in the simulation study. This could explain why the trimming method is less impactful.

The relative MSFE results of no trimming are seen in Table 8. For the one- and two-year horizon of the HICP and RGDP, the POET method results in a significant improvement compared to equal weights. For the one-year ahead forecast of UNEM, a significant improvement is seen as well. For the two-year ahead forecast of UNEM, no significant improvement is noted. Most notably, a 35% improvement for the one-year ahead forecast of UNEM. The forecast combination using a POET covariance estimate results in an improved forecast in 5 of the 6 settings.

Table 8: The relative MSFE values forecast combination without trimming methods. The weights have been estimated by the factor covariance matrix.

	1-year			2-year		
	HICP	RGDP	UNEM	HICP	RGDP	UNEM
MSFE	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)

Note: In order to assess whether the factor model method outperforms the standard equal weights method, the ratio of the mean squared forecast error (MSFE) using the factor covariance matrix and the MSFE of equal weights is examined, referred to as the relative MSFE. This table presents the results for these different two different forecast combination methods. The MSFE is averaged across the testing period. If this ratio is less than one, it indicates that the factor method outperforms the equal weights method. To test for statistical significance, the two-sided DM test is used. The p-values are shown between the brackets.

This is a clear indication that the forecast combination using the factor model outperforms the equal weights method in several settings while it hardly performs worse. The question remains, does combination using the POET model result in a better forecast combination than the data-driven trimming methods seen in Radchenko et al. (2023)? This is done by comparing the MSFE values of the POET combination with the MSFE values of the data-driven threshold method. The relative MSFE values of this evaluation are seen in Table 9. The MSFE values of the POET method are divided by the data-driven threshold method, so a value below 1 indicates a better performance of the POET method. The p-values are as before determined by the DM test and given in brackets.

Table 9: The relative MSFE values of the forecasting combination using the data-driven threshold method compared with the factor method. For the factor method, no trimming method is used.

	1-year			2-year		
$c_{min} = -2$	HICP	RGDP	UNEM	HICP	RGDP	UNEM
TR1	0.608(0.020)	0.997(0.950)	0.791(0.068)	0.702(0.014)	0.905(0.383)	1.305(0.104)
TR2	0.648(0.002)	0.960(0.428)	0.791(0.070)	0.700(0.015)	0.864(0.206)	1.232(0.281)
TR3	0.651(0.025)	0.981(0.616)	0.844(0.212)	0.679(0.009)	0.884(0.347)	1.385(0.061)
TR4	0.874(0.397)	1.001(0.984)	0.915(0.534)	0.833(0.074)	0.983(0.561)	1.134(0.207)
TR5	0.867(0.373)	1.011(0.810)	1.393(0.103)	0.802(0.025)	1.020(0.599)	1.254(0.044)

Note: In order to assess whether the factor model method outperforms the data-driven trimming threshold method, the ratio of the mean squared forecast error (MSFE) using factor covariance estimation method and the MSFE of the data-driven threshold method is examined, referred to as the relative MSFE. This table presents the results for these different two different forecast combination methods. The MSFE is averaged across the testing period. If this ratio is less than one, it indicates that the factor model method outperforms the data-driven trimming threshold method. To test for statistical significance, the two-sided DM test is used. The p-values are shown between the brackets.

For the one- and two-year ahead forecast for HICP the POET model results in an improved forecast combination compared to the data-driven model. The significance depends on the trimming method used. For the one- and two-year horizon for RGDP no significant difference is seen between the two methods. Lastly, for a one-year ahead forecast for UNEM the POET model results in an improved forecast combination compared to the data-driven model in some of the trimming methods. In the two-year ahead forecast, in most trimming methods no significant difference is seen. Only trimming methods 3 and 5 show a significantly improved performance of the data-driven method.

6 Conclusion

This thesis began with an extensive review of the current insights and challenges in forecast combination. The most significant challenge is the forecast combination puzzle, stating that the equal weights method frequently outperforms the optimal weight method. This is mainly due to estimation errors. Various ways to tackle this problem have been discussed. Firstly, the trimming method with a possible negative threshold proposed by Radchenko et al. (2023) has been explored. This method diminished the effect of the estimation errors. Secondly, the impact of several covariance estimation methods on estimation errors has been examined. Specifically, the sample covariance method, the linear shrinkage method and the factor method. The effect of covariance estimation methods on the forecast combination performance was evaluated through a simulation study and an implementation of an empirical application.

This thesis confirms the results regarding the trimming methods reported by Radchenko et al. (2023). Only the results from the trimming methods using constraint optimization differ from Radchenko et al. (2023) due to their stochastic components. Additionally, a coding mistake has been identified in determining the p-values for the data-driven threshold method. The conclusion remains the same. The trimming methods using an optimal trimming threshold, possibly below zero, allow for an optimal balance of the variance and bias. Thereby diminishing the effect of the estimation errors. This results in a forecast combination that outperforms equal weights in many settings.

In a simulation study, the linear shrinkage method and the factor model method lead to an improved forecasting combination compared to the standard sample covariance method. The difference, measured by the L2-norm, between the true covariance matrix and the estimated covariance matrix was decreased using these alternative methods. Indicating that fewer estimation errors are present compared to the sample covariance method. Interestingly, applying the trimming method in this simulation study did not improve the forecast combination. Only a performance deterioration is seen when the threshold comes close to the zero value. This indicates that the weights are less dispersed and that the bias resulting from the trimming is larger than the reduction in variance. When the covariance matrix is estimated by the linear shrinkage method and the factor model method, it is best to use the untrimmed optimal weights estimates, in contrast to what is determined by Radchenko et al. (2023). Once more, this confirms the importance of negative weights since trimming at zero is not optimal.

In the empirical application using these covariance estimation methods, it can be concluded that both alternative methods lead to an improved forecast combination compared to the equal weights method for most settings. Again the trimming methods did not lead to an improvement in the forecast combination. The untrimmed weights lead to the best-performing forecast combination, where negative weights were included. This is in agreement with the results of the simulation study. Additionally, both covariance estimation methods were compared to the data-driven threshold method developed by Radchenko et al. (2023). Using the alternative covariance estimation method led to an improved combination compared to the data-driven threshold in many settings.

To conclude, these findings highlight the importance of minimizing estimation errors at their root rather than solely mitigating their effects by trimming. The minimization of the estimation errors can be achieved by the use of the linear shrinkage method and the factor model method for covariance matrix estimation. Furthermore, this thesis confirms the relevance of negative weights in forecast combinations. More accurate and reliable forecasts can be generated by adopting improved covariance matrix estimation methods and incorporating negative weights.

Future research can explore additional covariance matrix estimation techniques and evaluate their performance in different forecast combination settings. The nonlinear shrinkage method could potentially improve the estimation of weights in settings where observations are not limited. Smith and Wallis (2009) discuss an estimation method where the covariances are set to zero, and only the variances are estimated. This could lead to fewer estimation errors a swell. Additionally, the performance of these methods can be explored by an empirical application involving the SPF of the Federal Reserve Bank of Philadelphia.

Acknowledgements

Throughout this thesis, I had the opportunity to explore two interconnected fields of forecast combination and covariance estimation. Exploring these areas' theoretical foundations and practical implications was fascinating and rewarding. I was delighted to see how these theoretical insights led to clear empirical improvements. This thesis has not only expanded my knowledge but has also made a valuable contribution to the estimation process of forecast combinations.

A Appendix

In this appendix, the derivations for the Monte Carlo Simulation seen in Radchenko et al. (2023) are presented. The paper does not explicitly provide the derivation of certain important characteristics of the forecast errors e_1 and e_2. These characteristics include the variance of the forecast errors, σ_1^2 and σ_2^2, as well as their covariance, cov(e_1, e_2)). Since the derivation of the true covariance matrix is essential for the extension, these characteristics will be derived for the purpose of replication.

First, forecast 1, y_1, for z_t is considered ,where $y_1 = z_{t-1}$ and z_t is generated from an AR(1) process. The forecast error is given by $e_1 = z_t - y_1$. The variance of the forecast error, denoted as σ_1^2, can be derived as follows:

$$\begin{aligned}
\sigma_1^2 &= \text{var}(e_1) \\
&= \text{var}(z_t - y_1) \\
&= \text{var}(z_t - z_{t-1}) \\
&= \text{var}(z_t) + \text{var}(-z_{t-1}) + 2\text{cov}(z_t, -z_{t-1}) \\
&= \sigma_z^2 + \sigma_z^2 - 2\text{cov}(z_t, z_{t-1}) \\
&= 2\sigma_z^2 - 2\text{cov}(\phi_1 z_{t-1} + e_t, z_{t-1}) \\
&= 2\sigma_z^2 - 2\phi_1\text{cov}(z_{t-1} + e_t, z_{t-1}) \\
&= 2\sigma_z^2 - 2\phi_1\text{cov}(z_{t-1}, z_{t-1}) \\
&= 2\sigma_z^2 - 2\phi_1\text{var}(z_{t-1}) \\
&= 2(1 - \phi_1)\sigma_z^2
\end{aligned}$$

In the derivation above, the definition of the forecast error is substituted in the first step. Then, the variance is rewritten using the formula var$(X+Y) =$ var$(X) +$ var$(Y) + 2$cov(X, Y). The time-invariant property of the AR(1) process is used, where both var(z_t) and var(z_{t-1}) have a variance of σ_z^2. In the covariance term, the definition of the AR(1) process is used: $z_t = \phi_1 z_{t-1} + e_t$. It is noted that the error term is independent and identically distributed standard-normal. As a result, this term drops out of the covariance since this means cov$(e_t, z_{t-1}) = 0$.

The same principles and steps can be applied to derive the σ_2^2 and cov(e_1, e_2). These derivations are presented without extensive commentary. Next, the variance of the second forecast for z_t will be derived. The second forecast is defined by: $y_2 = \rho_2 z_{t-2}$, where ρ_2 is the second autocorrelation term between the observations, $\rho_2 = $ cov$(z_t, z_{t-2}) = \phi_1^2$ with autoregressive parameter ϕ_1 from the AR(1) process. The forecast error is defined as by $e_2 = z_t - y_2$. This results in the following derivation:

$$\begin{aligned}
\sigma_2^2 &= \text{var}(e_2) \\
&= \text{var}(z_t - y_2) \\
&= \text{var}(z_t - \rho_2 z_{t-2}) \\
&= \text{var}(z_t) + \text{var}(-\rho_2 z_{t-1}) + 2\text{cov}(z_t, -\rho_2 z_{t-2}) \\
&= \sigma_z^2 + \rho_2^2 \text{var}(z_{t-1}) - 2\rho_2 \text{cov}(z_t, z_{t-2}) \\
&= \sigma_z^2 + \rho_2^2 \sigma_z^2 - 2\rho_2 \text{cov}(\phi_1 z_{t-1} + e_t, z_{t-2}) \\
&= \sigma_z^2 + \rho_2^2 \sigma_z^2 - 2\rho_2 \text{cov}(\phi_1(\phi_1 z_{t-2} + e_t) + e_t, z_{t-2}) \\
&= \sigma_z^2 + \rho_2^2 \sigma_z^2 - 2\rho_2 \text{cov}(\rho_2 z_{t-2} + \phi_1 e_{t-1} + e_t, z_{t-2}) \\
&= \sigma_z^2 + \rho_2^2 \sigma_z^2 - 2\rho_2^2 \text{cov}(z_{t-2}, z_{t-2}) \\
&= \sigma_z^2 + \rho_2^2 \sigma_z^2 - 2\rho_2^2 \sigma_z^2 \\
&= (1 - \rho_2^2)\sigma_z^2
\end{aligned}$$

Finally, the covariance between the two forecast error term will be determined, using the definitions and derivations mentioned before.

$$\begin{aligned}
\text{cov}(e_1, e_2) &= \text{cov}(z_t - z_{t-1}, z_t - \rho_2 z_{t-2}) \\
&= \text{cov}(z_t, z_t) + \text{cov}(z_t, -\rho_2 z_{t-2}) + \text{cov}(-z_{t-1}, z_t) + \text{cov}(-z_{t-1}, -\rho_2 z_{t-2}) \\
&= \sigma_z^2 - \rho_2 \text{cov}(z_t, z_{t-2}) - \text{cov}(z_{t-1}, z_t) + \rho_2 \text{cov}(z_{t-1}, z_{t-2}) \\
&= \sigma_z^2 - \rho_2 \text{cov}(\rho_2 z_{t-2} + \phi_1 e_{t-1} + e_t, z_{t-2}) - \text{cov}(z_{t-1}, \phi_1 z_{t-1} + e_t) \\
&+ \rho_2 \text{cov}(\phi_1 z_{t-2} + e_{t-1}, z_{t-2}) \\
&= \sigma_z^2 - \rho_2 \text{cov}(\rho_2 z_{t-2}, z_{t-2}) - \text{cov}(z_{t-1}, \phi_1 z_{t-1}) + \rho_2 \text{cov}(\phi_1 z_{t-2}, z_{t-2}) \\
&= \sigma_z^2 - \rho_2^2 \sigma_z^2 - \phi_1 \sigma_z^2 + \rho_2 \phi_1 \sigma_z^2 \\
&= \sigma_z^2 - \rho_2^2 \sigma_z^2 - \rho_1 \sigma_z^2 + \rho_2 \rho_1 \sigma_z^2 \\
&= \sigma_z^2 (1 - \rho_2^2 - \rho_1 + \rho_2 \rho_1) \\
&= \sigma_z^2 (1 - \rho_2)(1 - \rho_1 + \rho_2)
\end{aligned}$$

B Appendix

This appendix displays the results of using trimming methods 1,2, and 3 on the estimated weights in the empirical application. The results are seen in Table 10.

Table 10: The relative MSFE values of the forecast combination using trimming methods 1,2, and 3 with various trimming thresholds.

	1-year			2-year		
TR1	HICP	RGDP	UNEM	HICP	RGDP	UNEM
c: $-\infty$	8.959(0.139)	0.874(0.470)	1.065(0.870)	3.814(0.105)	5.339(0.333)	4.625(0.191)
-5.0	4.175(0.164)	0.878(0.487)	0.910(0.800)	2.320(0.059)	1.007(0.983)	4.625(0.191)
-4.5	4.155(0.167)	0.877(0.483)	0.910(0.800)	2.309(0.061)	1.015(0.964)	4.625(0.191)
-4.0	4.139(0.169)	0.876(0.478)	0.910(0.800)	2.296(0.064)	1.042(0.897)	4.625(0.191)
-3.5	4.129(0.171)	0.890(0.516)	0.910(0.799)	2.285(0.066)	1.039(0.903)	4.625(0.191)
-3.0	4.049(0.182)	0.904(0.566)	0.907(0.792)	2.276(0.068)	0.954(0.888)	4.594(0.194)
-2.5	4.005(0.188)	0.914(0.600)	0.903(0.784)	2.060(0.047)	0.938(0.851)	4.397(0.190)
-2.0	3.196(0.144)	0.870(0.297)	0.814(0.518)	1.742(0.061)	0.973(0.934)	2.190(0.067)
-1.5	2.197(0.065)	0.823(0.055)	0.644(0.134)	1.552(0.121)	0.938(0.811)	1.808(0.169)
-1.0	1.655(0.086)	0.821(0.042)	0.555(0.027)	1.416(0.223)	0.843(0.199)	1.436(0.352)
-0.5	1.196(0.315)	0.852(0.038)	0.594(0.002)	1.129(0.439)	0.900(0.085)	0.735(0.332)
0.0	1.021(0.554)	0.972(0.004)	0.914(0.030)	1.026(0.418)	0.968(0.178)	0.804(0.004)
TR2						
c: $-\infty$	8.959(0.139)	0.874(0.470)	1.065(0.870)	3.814(0.105)	5.339(0.333)	4.625(0.191)
-5.0	6.456(0.092)	0.901(0.578)	0.903(0.784)	2.803(0.042)	1.533(0.398)	4.625(0.191)
-4.5	7.364(0.109)	0.914(0.636)	0.903(0.785)	2.729(0.039)	1.399(0.441)	4.625(0.191)
-4.0	6.720(0.097)	0.941(0.759)	0.904(0.787)	2.773(0.041)	1.292(0.505)	4.625(0.191)
-3.5	7.185(0.103)	0.937(0.736)	1.227(0.605)	2.685(0.039)	1.231(0.563)	4.625(0.191)
-3.0	6.795(0.100)	0.931(0.705)	2.624(0.375)	2.595(0.039)	1.257(0.557)	4.615(0.192)
-2.5	6.044(0.093)	0.925(0.675)	2.571(0.377)	2.473(0.037)	1.107(0.768)	4.547(0.192)
-2.0	5.331(0.087)	0.911(0.589)	2.017(0.417)	2.302(0.032)	0.994(0.985)	3.727(0.152)
-1.5	4.275(0.070)	0.873(0.360)	1.606(0.474)	2.124(0.036)	0.891(0.728)	3.326(0.136)
-1.0	3.203(0.060)	0.849(0.226)	0.852(0.681)	1.898(0.044)	0.802(0.435)	2.174(0.106)
-0.5	2.205(0.050)	0.871(0.169)	0.547(0.016)	1.431(0.111)	0.778(0.169)	1.251(0.501)
0.0	1.021(0.554)	0.972(0.004)	0.914(0.030)	1.026(0.418)	0.968(0.178)	0.804(0.004)
TR3						
c: $-\infty$	8.959(0.139)	0.874(0.470)	1.065(0.870)	3.814(0.105)	5.339(0.333)	4.625(0.191)
-5.0	5.681(0.087)	0.874(0.472)	0.955(0.902)	2.803(0.042)	1.002(0.996)	4.625(0.191)
-4.5	5.371(0.089)	0.874(0.470)	0.950(0.891)	2.729(0.039)	0.990(0.977)	4.625(0.191)
-4.0	5.130(0.094)	0.872(0.466)	0.946(0.880)	2.647(0.038)	0.984(0.961)	4.625(0.191)
-3.5	4.890(0.103)	0.876(0.474)	0.936(0.858)	2.581(0.039)	0.977(0.945)	4.625(0.191)
-3.0	4.669(0.116)	0.881(0.488)	0.921(0.824)	2.521(0.040)	0.933(0.840)	4.615(0.192)
-2.5	4.475(0.132)	0.887(0.507)	0.914(0.809)	2.396(0.037)	0.894(0.749)	4.547(0.192)
-2.0	4.018(0.131)	0.877(0.427)	0.872(0.695)	2.168(0.035)	0.878(0.716)	3.537(0.144)
-1.5	3.220(0.099)	0.856(0.258)	0.708(0.249)	1.921(0.041)	0.865(0.663)	2.657(0.084)
-1.0	2.393(0.077)	0.841(0.128)	0.568(0.052)	1.664(0.082)	0.810(0.371)	1.921(0.098)
-0.5	1.552(0.120)	0.850(0.058)	0.490(0.002)	1.317(0.182)	0.841(0.174)	0.854(0.587)
0.0	1.021(0.554)	0.972(0.004)	0.914(0.030)	1.026(0.418)	0.968(0.178)	0.804(0.004)

Note: To assess whether the trimming method outperforms the standard equal weights method, the ratio of the mean squared forecast error (MSFE) using trimmed weights and the MSFE of equal weights is examined, referred to as the relative MSFE. This table presents the results for these different two different forecast combination methods for different trimming thresholds c, indicated on the left side. The MSFE is averaged across the testing period. If this ratio is less than one, it indicates that the trimming method outperforms the equal weights method. To test for statistical significance, the two-sided DM test is used. The p-values are shown between the brackets.

Table 11 shows the optimal trimming thresholds for each testing period across all trimming methods for the one-year ahead forecast of UNEM and the one-year ahead forecast of HICP. These trimming thresholds are determined by a pseudo out-of-sample evaluation, with $c_{min} = -2$.

Table 11: The optimal c values for all test periods in the data-driven threshold method for the one-year ahead forecasts for HICP and UNEM.

HICP 1-year					
Test period	TR1	TR2	TR3	TR4	TR5
1	0	0	0	0	-0.1
2	0	0	0	0	0
3	-0.4	-0.1	-0.2	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	-0.5	-0.1	0	0	0
7	0	0	-0.1	0	0
8	-0.1	0	-0.1	0	0
9	0	0	0	0	0
10	-0.5	0	-0.2	0	0
11	0	0	0	0	0
12	-0.2	-0.1	-0.1	0	0
13	0	0	0	0	0
14	0	0	0	0	0
15	-0.7	-0.2	-0.3	0	0
16	-0.7	-0.2	-0.5	0	0
UNEM 1-year					
Test period					
1	-0.4	-0.2	-0.3	0	-0.1
2	-0.3	-0.1	-0.3	0	0
3	-1.6	0	-0.3	0	-0.5
4	-0.3	-0.1	-0.3	0	0
5	0	0	-0.4	0	-0.1
6	-1.4	-0.4	-0.4	0	0
7	0	0	0	0	-0.1
8	-0.1	0	-0.2	0	0
9	-0.4	-0.2	-0.3	0	0
10	-1	-0.1	-0.3	0	-0.1
11	-0.2	-0.1	-0.2	0	-0.1
12	-0.3	-0.1	-0.2	0	-0.1
13	-0.3	-0.2	-0.2	0	0
14	-0.4	-0.1	-0.3	0	0
15	-0.3	-0.2	-0.2	0	-0.1
16	-0.5	-0.2	-0.3	0	0

In Figure 6 a histogram is presented of the weights obtained from trimming method 3 of the one-year ahead forecast of HICP and the histogram of the one-year ahead forecast of UNEM across the 16 testing periods.

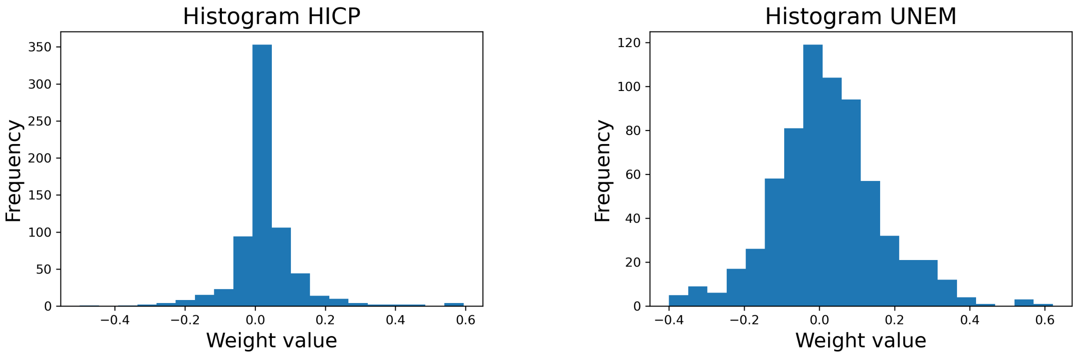

Figure 6: The histograms of the weights after using trimming method 3 in the data-driven trimming method. On the right is the histogram of the weights for the forecast combination of the one-year ahead forecasts of the HICP. On the left is the histogram of the weights for the forecast combination of the one-year ahead forecasts of the UNEM.

C Appendix

In this Appendix the characteristics of the forecast errors in this high-dimensional setting are derived. In Appendix A, the important characteristics of the forecast errors were derived in the two forecast setting of Radchenko et al. (2023). To evaluate different covariance estimation methods, a Monte Carlo simulation is performed in a high-dimensional setting. To achieve the high-dimensional setting, additional forecasts are necessary.
As discussed in Subsection 5.1.1, the auto-correlation coefficient function ρ_i is defined as the correlation between observations z_t and z_{t-i}, given by $\rho_i = \text{corr}(z_t, z_{t-i}) = \phi_1^i$, with ϕ_1 being the autoregressive parameter of the AR(1) process. The 25 forecasts for z_t are generated by:

$$y_1 = z_{t-1} \qquad\qquad y_i = \rho_i z_{t-i}$$

for $i \in 2, \ldots, 25$. The forecast error for the first forecast, e_{1t}, is defined as $e_{1t} = z_t - z_{t-1}$. The forecast errors for the additional forecasts, e_{it}, are defined as $e_{it} = z_t - \rho_i$, where $i \in 2, \ldots, 25$. Since forecasts 2 to 25 have the same unbiased property, they also have a similar variance structure for the forecast errors. The variance of forecast error e_1 is still the same, $\sigma_1^2 = 2(1-\phi_1)\sigma_z^2$. The variance of the forecast error e_i can be calculated as follows:

$$\begin{aligned}
\sigma_i^2 &= \text{var}(e_i) \\
&= \text{var}(z_t - \rho_i z_{t-i}) \\
&= \text{var}(z_t) + \text{var}(-\rho_i z_{t-i}) + 2\text{cov}(z_t, -\rho_i z_{t-i}) \\
&= \sigma_z^2 + \rho_i^2 \text{var}(z_{t-1}) - 2\rho_i \text{cov}(z_t, z_{t-i}) \\
&= \sigma_z^2 + \rho_i^2 \text{var}(z_{t-1}) - 2\rho_i \text{cov}(\rho_i z_{t-i}, z_{t-i}) \\
&= (1-\rho_i^2)\sigma_z^2
\end{aligned}$$

The derivation is shorted since the structure is similar to the derivation of σ_2^2. Since there are two types of forecasts, y_1 and y_i for $i \in 2, \ldots, 25$, there are two formulas to determine the true covariance. The first formula calculates the covariance between forecast error e_1 and e_i, given by:

$$\begin{aligned}
\text{cov}(e_1, e_i) &= \text{cov}(z_t - z_{t-1}, z_t - \rho_i z_{t-i}) \\
&= \text{cov}(z_t, z_t) + \text{cov}(z_t, -\rho_i z_{t-i}) + \text{cov}(-z_{t-1}, z_t) + \text{cov}(-z_{t-1}, -\rho_i z_{t-i}) \\
&= \sigma_z^2 - \rho_i \text{cov}(z_t, z_{t-i}) - \rho_1 \sigma_z^2 + \rho_i \text{cov}(z_{t-1}, z_{t-i}) \\
&= \sigma_z^2 - \rho_1 \sigma_z^2 - \rho_i \text{cov}(z_t, z_{t-i}) + \rho_i \text{cov}(z_{t-1}, z_{t-i}) \\
&= \sigma_z^2 - \rho_1 \sigma_z^2 - \rho_i \text{cov}(\rho_i z_{t-i}, z_{t-i}) + \rho_i \text{cov}(\rho_{i-1} z_{t-i}, z_{t-i}) \\
&= \sigma_z^2 - \rho_1 \sigma_z^2 - \rho_i \rho_i \sigma_z^2 + \rho_i \rho_{i-1} \sigma_z^2 \\
&= \sigma_z^2 (1 - \rho_1 - \rho_i^2 + \rho_i \rho_{i-1}) \\
&= \sigma_z^2 (1 - \rho_1 - \rho_i^2 + \rho_i \rho_{i-1})
\end{aligned}$$

The second formula calculates the covariance between the error of two unbiased forecasts, where $i, j \geq 2$ and $i \leq j$, given by:

$$\begin{aligned}
\text{cov}(e_i, e_j) &= \text{cov}(z_t - \rho_i z_{t-i}, z_t - \rho_j z_{t-j}) \\
&= \text{cov}(z_t, z_t) + \text{cov}(z_t, -\rho_i z_{t-i}) + \text{cov}(-\rho_i z_{t-i}, z_t) + \text{cov}(-\rho_i z_{t-i}, z_t - \rho_j z_{t-j}) \\
&= \sigma_z^2 - \rho_i \text{cov}(z_t, z_{t-i}) - \rho_j \text{cov}(z_t, z_{t-j}) + \rho_i \rho_j \text{cov}(z_{t-i}, z_{t-j}) \\
&= \sigma_z^2 - \rho_i^2 \sigma_z^2 - \rho_j^2 \sigma_z^2 + \rho_i \rho_j \text{cov}(\phi_1^{j-i} z_{t-j}, z_{t-j}) \\
&= \sigma_z^2 - \rho_i^2 \sigma_z^2 - \rho_j^2 \sigma_z^2 + \rho_i \rho_j \rho_{j-i} \text{cov}(z_{t-j}, z_{t-j}) \\
&= \sigma_z^2 - \rho_i^2 \sigma_z^2 - \rho_j^2 \sigma_z^2 + \rho_i \rho_j \rho_{j-i} \sigma_z^2 \\
&= \sigma_z^2 (1 - \rho_i^2 - \rho_j^2 + \rho_i \rho_j \rho_{j-i})
\end{aligned}$$

With these formulas, all tools are available to determine the true covariance matrix.

D Appendix

In this Appendix the additional results from the Monte Carlo Simulation in high-dimension setting are denoted. In Appendix C, the tools are derived to obtain the true covariance matrix given a number of forecasts n and the autoregressive parameter ϕ_1 for the AR(1) process. In the Monte Carlo simulation, the number of forecasts is set at 25 and $\phi_1 = -0.95$. Given the true covariance, the optimal weights can be determined. This results in the optimal weight vector w^*:

$$\begin{aligned}
w^* =&[0.00197, 0.95653, 0.08967, -0.09439, 0.09936, \\
&-0.10459, 0.11009, -0.11588, 0.12198, -0.12840, \\
&0.13516, -0.14228, 0.14976, -0.15765, 0.16594, \\
&-0.17468, 0.18387, -0.19355, 0.20373, -0.21446, \\
&0.22574, -0.23763, 0.25013, -0.26330, 0.13284]'
\end{aligned}$$

In the optimal weight vector, 11 of the 25 forecasts have a negative weight optimal weight. The most negative optimal weight is -0.26.

E Appendix

The relative MSFE values of the data-driven shrinkage methods are determined for trimming methods 1,2 and 3. The results are seen in Table 12.

Table 12: The relative MSFE values of the forecast combination using the trimming methods 1,2 and 3 with various trimming thresholds. The weights have been estimated by the data-driven linear shrinkage covariance matrix.

	1-year			2-year		
TR1	HICP	RGDP	UNEM	HICP	RGDP	UNEM
c: $-\infty$	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-5.0	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-4.5	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-4.0	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-3.5	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-3.0	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-2.5	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-2.0	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-1.5	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-1.0	0.985(0.660)	0.928(0.014)	0.380(0.000)	1.027(0.543)	0.798(0.000)	0.560(0.006)
-0.5	1.194(0.127)	0.928(0.014)	0.401(0.000)	0.987(0.728)	0.798(0.000)	0.556(0.006)
0.0	0.967(0.267)	0.961(0.003)	0.803(0.000)	0.957(0.086)	0.912(0.002)	0.894, (0.052)
TR2						
c: $-\infty$	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-5.0	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-4.5	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-4.0	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-3.5	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-3.0	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-2.5	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-2.0	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-1.5	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-1.0	0.985(0.660)	0.928(0.014)	0.380(0.000)	1.027(0.543)	0.798(0.000)	0.560(0.006)
-0.5	0.985(0.660)	0.928(0.014)	0.391(0.000)	0.978(0.584)	0.798(0.000)	0.560(0.006)
0.0	0.967(0.267)	0.961(0.003)	0.803(0.000)	0.957(0.086)	0.912(0.002)	0.894(0.052)
TR3						
c: $-\infty$	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-5.0	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-4.5	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-4.0	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-3.5	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-3.0	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-2.5	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-2.0	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-1.5	0.985(0.660)	0.928(0.014)	0.388(0.000)	1.009(0.740)	0.798(0.000)	0.560(0.006)
-1.0	0.985(0.660)	0.928(0.014)	0.380(0.000)	1.027(0.543)	0.798(0.000)	0.560(0.006)
-0.5	1.145(0.257)	0.928(0.014)	0.373(0.000)	1.051(0.437)	0.798(0.000)	0.552(0.005)
0.0	0.967(0.267)	0.961(0.003)	0.803(0.000)	0.957(0.086)	0.912(0.002)	0.894(0.052)

Note: To assess whether the trimming method based on the linear shrinkage parameter outperforms the standard equal weights method, the ratio of the MSFE using trimmed weights and the MSFE of equal weights is examined, referred to as the relative MSFE. This table presents the results for different trimming thresholds c. The MSFE is averaged across the testing period. If this ratio is less than one, it indicates that the trimming method outperforms the equal weights method. To test for statistical significance, the two-sided DM test is used. The p-values are shown between the brackets.

The relative MSFE values of the factor model methods is determined for trimming methods 1,2 and 3. The results are seen in Table 13.

Table 13: The relative MSFE values of the forecast combination using the trimming methods 1,2 and 3 with various trimming thresholds. The weights have been estimated by the factor model covariance matrix.

	1-year			2-year		
TR1	HICP	RGDP	UNEM	HICP	RGDP	UNEM
c: $-\infty$	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.50	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.45	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.40	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.35	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.30	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.25	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.797(0.000)	1.085(0.511)
-0.20	0.756(0.091)	0.894(0.001)	0.644(0.000)	0.731(0.016)	0.798(0.000)	1.085(0.511)
-0.15	0.758(0.092)	0.895(0.001)	0.655(0.000)	0.739(0.016)	0.802(0.000)	1.084(0.514)
-0.10	0.777(0.074)	0.903(0.001)	0.684(0.000)	0.763(0.013)	0.810(0.000)	1.069(0.567)
-0.05	0.839(0.058)	0.921(0.001)	0.742(0.000)	0.818(0.011)	0.828(0.000)	1.057(0.583)
0.00	0.937(0.063)	0.952(0.000)	0.847(0.000)	0.897(0.013)	0.879(0.000)	1.022(0.765)
TR2						
c: $-\infty$	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.50	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.45	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.40	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.35	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.30	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.25	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.797(0.000)	1.085(0.511)
-0.20	0.756(0.091)	0.894(0.001)	0.643(0.000)	0.730(0.016)	0.798(0.000)	1.085(0.511)
-0.15	0.758(0.091)	0.895(0.001)	0.653(0.000)	0.736(0.017)	0.801(0.000)	1.084(0.514)
-0.10	0.772(0.076)	0.901(0.001)	0.676(0.000)	0.755(0.014)	0.808(0.000)	1.072(0.556)
-0.05	0.815(0.061)	0.915(0.001)	0.719(0.000)	0.797(0.011)	0.822(0.000)	1.060(0.572)
0.00	0.937(0.063)	0.952(0.000)	0.847(0.000)	0.897(0.013)	0.879(0.000)	1.022(0.765)
TR3						
c: $-\infty$	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.50	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.45	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.40	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.35	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.30	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.796(0.000)	1.085(0.511)
-0.25	0.756(0.091)	0.894(0.001)	0.642(0.000)	0.728(0.016)	0.797(0.000)	1.085(0.511)
-0.20	0.756(0.091)	0.894(0.001)	0.643(0.000)	0.730(0.016)	0.798(0.000)	1.085(0.511)
-0.15	0.758(0.091)	0.895(0.001)	0.653(0.000)	0.744(0.016)	0.801(0.000)	1.084(0.514)
-0.10	0.777(0.072)	0.904(0.001)	0.692(0.000)	0.767(0.013)	0.810(0.000)	1.068(0.570)
-0.05	0.854(0.059)	0.926(0.001)	0.762(0.000)	0.830(0.013)	0.837(0.000)	1.061(0.559)
0.00	0.937(0.063)	0.952(0.000)	0.847(0.000)	0.897(0.013)	0.879(0.000)	1.022(0.765)

Note: To assess whether the trimming method based on the factor model outperforms the standard equal weights method, the ratio of the mean squared forecast error (MSFE) using trimmed weights and the MSFE of equal weights is examined, referred to as the relative MSFE. This table presents the results for different trimming thresholds c. The MSFE is averaged across the testing period. If this ratio is less than one, it indicates that the trimming method outperforms the equal weights method. To test for statistical significance, the two-sided DM test is used. The p-values are shown between the brackets.

References

Bates, J. M. and C. W. J. Granger (1969). "The Combination of Forecasts". In: *Journal of the Operational Research Society* 20, pp. 451–468.

Claeskens, G., J. R. Magnus, A. L. Vasnev, and W. Wang (2016). "The forecast combination puzzle: A simple theoretical explanation." In: *International journal of Forecasting* 32, pp. 754–762.

Diebold, F. and R. Mariano (1995). "Comparing Predictive Accuracy." In: *Journal of Business and Economic Statistics* 13.3, pp. 134–144.

Fan, J., Y. Liao, and M. Mincheva (2013). "Large Covariance Estimation by Thresholding Principal Orthogonal Complements." In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75, pp. 603–680.

Fan, J., j. Zhang, and K. Yu (2012). "Vast portfolio selection with gross-exposure constraints." In: *Journal of the American Statistical Association* 107.498, pp. 592–606.

James, W. and C. Stein (1961). "Estimation with Quadratic Loss". In: *University of California Press* 1, pp. 361–380.

Ledoit, O. and M. Wolf (2004). "A well-conditioned estimator for large-dimensional covariance matrices." In: *Journal of Multivariate Analysis* 88.2, pp. 365–411.

— (2021). "Shrinkage estimation of large covariance matrices: Keep it simple, statistician?" In: *Journal of Multivariate Analysis* 186, p. 104796.

Magnus, J. R. and G. De Luca (2016). "Weighted-average least squared (WALS): A survey." In: *Journal of Economic Surveys* 30, pp. 117–148.

Magnus, J.R. (1999). "The traditional pretest estimator". In: *Theory of Probability and its Applications* 44, pp. 293–308.

Matsypura, D., R. Thompson, and A. L. Vasnev (2018). "Optimal selection of expert forecasts with integer programming". In: *Omega* 78, pp. 165–175.

Nard, G. De, O. Ledoit, and M. Wolf (2021). "Factor Models for Portfolio Selection in Large Dimensions: The Good, the Better and the Ugly." In: *Journal of Financial Econometrics* 19.2, pp. 236–257.

Newbold, P. and C. W. J. Granger (1974). "Experience with Forecasting Univariate Time Series and the Combination of Forecasts." In: *Journal of the Royal Statistical Society* 137.2, pp. 131–165.

Radchenko, P., A. Vasnev, and W. Wang (2023). "Too similar to combine? On negative weights in forecast combination". In: *International Journal of Forecasting* 39.1, pp. 18–38.

Ross, S. (1976). "The Arbitrage Theory of Capital Asset Pricing." In: *Journal of Economic Theory* 13, pp. 341–360.

Smith, J. and K. Wallis (2009). "A Simple Explanation of the Forecast Combination Puzzle". In: *Oxford Bulletin of Economics and Statistics* 71, pp. 331–355.