Katholische Universität Eichstätt-Ingolstadt

Summary and abstracts of the cumulative dissertation
presented for the attainment of the degree of
Doctor rerum politicarum (Dr. rer. pol.)

Integrated Models for the Selection and Weighting of Individual Opinions and Forecasts in Expectation and Forecast Combination

Integrierte Modelle zur Selektion und Gewichtung individueller Meinungen und Vorhersagen in der Erwartungs- und Prognosekombination

Chair for Business Administration and Business Informatics
Ingolstadt School of Management
Katholische Universität Eichstätt-Ingolstadt

proposed by
Felix Schulz, M. Sc.

First examiner
Prof. Dr. Thomas Setzer

Second examiner
Prof. Dr. Heinrich Kuhn

Day of the oral examination
20.06.2023

Ingolstadt, 2023

Overview

The present cumulative dissertation was written at the Chair for Business Administration and Business Informatics at the Faculty of Business and Economics at the Katholische Universität Eichstätt-Ingolstadt and is concerned with the development, application, and calibration of integrated models for the selection and weighting of individual opinions and predictions in expectation and forecast combination. The following four research contributions were made in the context of the PhD:

1. Schulz, Felix; Setzer, Thomas; Balla, Nathalie (2022): **Linear Hybrid Shrinkage of Weights for Forecast Selection and Combination**, in: Proceedings of the 55th Hawaii International Conference on System Sciences [VHB-JQ3: C, CORE2018: A].

2. Schulz, Felix (2022): **Non-Linear Hybrid Shrinkage of Weights for Forecast Selection and Combination**, in: Wirtschaftsinformatik 2022 Proceedings. 7 [VHB-JQ3: C, CORE2018: C].

3. Schulz, Felix; Setzer, Thomas; Balla, Nathalie (2023): **A Combined Measure Based on Diversification and Accuracy Gains for Forecast Selection in Forecast Combination**, in: Operations Research Proceedings 2022 (forthcoming) [VHB-JQ3: D, CORE2018: C].

4. Schulz, Felix; Setzer, Thomas (2023): **Shrinkage of Weights Towards Subset Selection in Forecast Combination**. Working Paper.

   The published or submitted versions of these contributions may differ slightly from the versions in this paper for consistency reasons. This does not affect the content of accepted papers. The content of working papers, on the other hand, may still change during the review process.

In forecast combination, multiple predictions are linearly combined through the assignment of weights to individual forecast models or forecasters. Various approaches exist for defining the weights, which typically involve determining the number of models to be used for combination (selection), choosing an appropriate weighting function for the forecast scenario (weighting), and using regularization techniques to adjust the calculated weights (shrinkage). The papers listed address the integration of the three approaches into holistic data analytical models. The first paper develops a two-stage model in which weights are first calculated based on the in-sample error covariances to minimize the error on the available data by combining the individual models. Based on the selection status of a model, the weight of an individual model is then linearly shrunk either toward the mean or toward zero. The selection status is thereby derived apriori from information criteria, where Contribution 1 introduces the selection based on the model's in-sample accuracy and performance robustness under uncertainty. Contribution 2 modifies the two-stage model to shrink the forecasters' weights non-proportionally to the mean or zero. Further, a new information criterion based on forward feature selection is proposed that iteratively selects the forecaster that is expected to achieve the largest increase in accuracy when combined. Contribution 3 extends the iteration-based information criterion presented in the second contribution to consider diversity gains in addition to accuracy gains when selecting and combining forecasters. A one-stage model for simultaneous weighting, shrinking, and selection is finally built and evaluated on simulated data in Contribution 4. Instead of requiring a prior selection criterion, the model itself learns which forecasters to shrink to the mean or to zero, while relying on a new sampling procedure to tune the model.

Contents

List of Tables

List of Abbreviations

eLASSO Egalitarian LASSO

HS Hybrid Shrink

LASSO Least Absolute Shrinkage and Selection Operator

LHS Linear Hybrid Shrinkage

MSE Mean Squared Error

NLHS Non-Linear Hybrid Shrinkage

OW Optimal Weights

peLASSO Partially-egalitarian LASSO

SA Simple Average

SubA Subset Averaging

1 Motivation and Aim

Forecasts represent an essential part of the decision-making and planning processes of companies and economies. At the economic level, for example, forecasts provide information on the development of macroeconomic indicators such as GDP growth, inflation, and interest rates (Banerjee et al., 2005). In the corporate context, forecasts support areas like sales, production, and financial planning with critical input (Granger, 2014). There is ample evidence that combining forecasts leads to higher accuracy compared to using a single forecast, as shown in various scientific fields (e.g. Diebold & Pauly, 1990; Rapach et al., 2010) and in representative forecasting competitions (e.g. Makridakis et al., 2018; Bojer & Meldgaard, 2021; Makridakis et al., 2022).

In forecast combination, a set of forecasts is linearly combined into a weighted forecast, thereby incorporating different information and assumptions of the underlying models and experts (Bates & Granger, 1969). The most known forecast combination models are called *Optimal Weights* (OW) (Bates & Granger, 1969) and *Simple Average* (SA) (Makridakis & Winkler, 1983). Both weighting approaches are of particular interest in terms of the bias–variance trade-off known from statistical learning theory (e.g., Smith & Wallis, 2009; Atiya, 2020; Blanc & Setzer, 2020). OW exploit the information from the underlying error data of the individual models and learn weights to reduce the error in the sample (bias) to the smallest possible level. SA, on the other hand, stoically assigns equal weights to the models regardless the available data, but reduces the cross-sample error (variance) by fixing weights.

Due to the systematic under- and overfitting of the combined forecasts compared to the actual outcome by choosing one of the two weighting approaches, so-called forecast shrinkage models try to form a balance between the bias-minimizing OW and the variance-minimizing SA (Blanc & Setzer, 2020). By using a shrinkage parameter, this approach forms a (non-)linear weight combination of OW and SA, where each step toward OW represents an information gain for the combined forecast by including more individual forecast information, while each step toward SA promotes equality among forecasters and represents the safer choice under enormous uncertainty in the learning process.

The simple shrinkage of OW to SA neglects the possibility of direct selection of forecasters. Forecast selection is another approach to combine forecasts and describes the task of selecting or excluding forecasters from the set of available forecasters, whereby the weights of the selected forecasters are usually averaged, known as *Subset Averaging* (SubA). Forecast selection has been successfully tested in the area of point forecasts (Aiolfi & Timmermann, 2006), probability forecasts (Geweke & Amisano, 2011), and judgmental forecasts (Mannes et al., 2014).

The advantages of averaging a subset while excluding the remaining forecasts are closely related to those of SA. Except for the selection step, SubA is an easy model to implement. Once the final selection is made, SubA minimizes the variance component by fixing weights. Their difference is that more weight is assigned to the selected forecasters than in SA, while potentially worse forecasters are excluded from the selection. Although in this way a larger amount of information is included in the combination compared to SA, all (non-)selected forecasters receive the same weight, neglecting individual information from forecasters with importance to the combined forecast. An intuitive approach for learning weights closer to their actual weight vectors would be to combine forecast selection by shrinking OW toward SubA, thereby relying on SubA only in the case of large uncertainty.

The main research question of this thesis is to evaluate whether hybrid shrinkage of weights toward the two targets of mean and zero, i.e., toward subset average weight combinations, can improve out-of-sample forecast accuracy. To answer the research question, a series of mathematical-statistical models are developed, applied, and evaluated on simulated data that integrate simultaneous weighting, shrinkage, and selection into a holistic model by learning OW and shrinking them toward SubA, while flexibly averaging and excluding forecasters from the combination along the shrinkage path. Within the scope of this work is to derive helpful decision criteria to provide guidance for the choice between integrated models and models that focus on only one of the tasks of weighting, shrinkage, and selection. For this reason, the proposed integrated models are evaluated exclusively on simulated data that allow to systematically change various underlying conditions known in the forecasting area such as the number of available forecast models, the number of available training data, the variances of the forecasters and the correlations of the individual forecast errors. For the simulation process, the error data are drawn from a multivariate normal distribution with zero mean, which implies that individual efficient point forecasts are assumed.

2 Outline of the Cumulative Dissertation

In the following thesis, integrated models for simultaneous weighting, shrinkage, and selection in forecast combination are presented, which are summarized in Figure 1. The family of integrated models for the combination of forecasts is divided into two-stage and one-stage models, where in two-stage models the forecasters are first selected according to information criteria and in the second stage shrunk against the mean or zero. In the first three papers, four information criteria for the selection of forecasters and two formulas for the subsequent weighting and shrinkage are proposed, whereby all criteria and formulas can be applied interchangeably. The main contribution of this work is the development of a one-stage model that inherits the forecast selection in a constrained optimization function, which is presented in the fourth paper. In the following, the individual research contributions are briefly presented in chronological order.

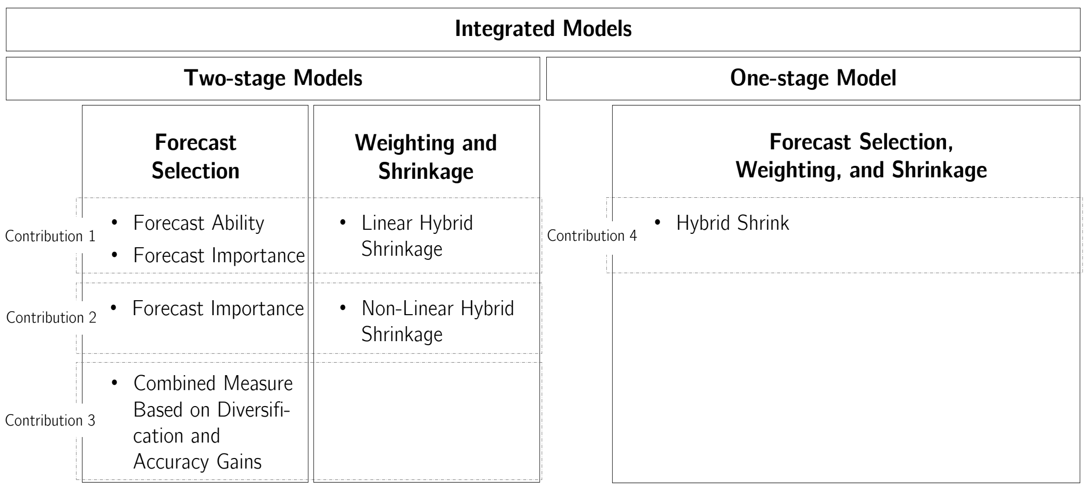

Figure 1: Overview of the contributions to the cumulative dissertation

1. Research Contribution

Following the central research question of the dissertation, the first paper describes the development of a model for hybrid shrinkage of forecast weights toward zero and the mean. The design of the model is inspired by the *Partially-egalitarian LASSO* (peLASSO) presented by Diebold & Shin (2019). peLASSO is a two-stage model that first selects the best individual forecasters by shrinking OW of all forecasters toward zero using the *Least Absolute Shrinkage and Selection Operator* (LASSO). In the second stage, the weights of the survivors are then either directly averaged or re-learned using OW and shrunk to the mean. Although peLASSO represents a novelty in forecast combination by integrating forecast selection, weighting, and shrinkage in just one model, the selection criterion and model design of peLASSO motivate the development of an alternative approach.

Considering that predictions in practice are often highly correlated due to similar underlying information, forecast selection via LASSO could often be suboptimal, as LASSO tends to randomly select one variable from a group of correlated variables (Zou & Hastie, 2005). In addition to the used selection criterion, the design of peLASSO by directly deleting forecasters is debatable. Although reducing the pool of forecasters may give more importance to selected forecasters when relearning weights, the approach neglects the possibility that excluded forecasters may still provide useful information for the combined forecast in the future. Finally, the results of peLASSO turn out to be very sensitive to its hyperparameters learned during cross-validation, which raises questions about its practicability. As an alternative, the authors even propose a simpler model to perform SubA that is intended to mimic the observed behavior of peLASSO, but proves to be very computationally intensive.

Based on the above considerations, an alternative two-stage forecast combination model called *Linear Hybrid Shrinkage* (LHS) is presented. Using a shrinkage parameter, LHS forms a linear combination of OW and SubA, where the number of equally weighted forecasters for SubA can be flexibly set as a hyperparameter. LHS does not directly exclude forecasters, but shrinks their OW toward zero or the mean, so that selection above zero occurs only at maximum shrinkage. To determine which forecasters are shrunk in which direction, information criteria from statistical learning theory are used, with one being a simple criterion for selection based on the forecasters' in-sample forecast performance and the other being a criterion for considering interactions between forecasts based on permutation-based variable importance. For the latter, uncertainty is introduced by repeatedly and randomly shuffling the validation data in cross-validation and assessing the robustness of the results based on their performance change derived from learned OW.

To evaluate LHS compared to other forecast combination models, an analysis based on the classification of simulation results is performed. First, the best model in terms of out-of-sample Mean Squared Error (MSE) for each combination of simulation parameters is determined and set as the target variable. Then, a decision tree is learned using the simulation parameters of training size, variance, correlation, and number of forecasters as predictors and the best model as predictand.

The classification results show that at high pairwise correlations between forecasters LHS is mainly dominated by OW and non-linear shrinkage of OW towards SA. An explanation for this observation is that LHS has difficulty regulating the widely dispersed OW in high correlation settings due to its linear shrinkage behavior. In contrast, the greatest advantage of the two-target shrinkage by LHS is evident when correlations are low to medium and the number of forecasters is twelve or more.

2. Research Contribution

While the first paper focused on a linear shrinkage of OW toward SubA, the second contribution presents a non-linear method to realize possible further accuracy improvements on unseen data by hybrid shrinkage from OW to SubA. The motivation for introducing a non-linear shrinkage method lies in the OW, which are used as a baseline for hybrid shrinkage.

In scenarios of increased uncertainty with a small number of available training data relative to the number of forecasts, OW can be massively overestimated. To balance the extreme weights of one or more forecasters, LHS can solely shrink the weights of all forecasters globally to the mean or zero. The increasing shrinkage of all forecasters, however, may result in the removal of most of the information generated from the training set. By introducing a non-linear formulation of hybrid shrinkage, the model itself can learn which forecasters and to which extent an individual regularization of the forecaster weights is required. With this gained flexibility, the extreme weights of one or more forecasts can be regularized more strongly to zero or the mean, while the weights of other forecasts can remain unaffected by the shrinkage.

The respective model, termed *Non-Linear Hybrid Shrinkage* (NLHS), relies on an OLS term to derive OW and a LASSO-based regularization term to penalize weight deviations from a given selection vector. The vector, thereby, contains per forecaster either zero or equal weights depending on the forecasters' individual selection status. As with LHS, NLHS is a two-stage integrated forecast combination method with a prior forecast selection step. For forecast selection, again selection based on the forecaster's in-sample performance is used, while another information criterion for incorporating interactions between forecasts is proposed on the basis of forward feature selection. The new criterion represents an iterative process that starts with the selection of the best forecaster in the sample and gradually adds other forecasters to the selection based on the resulting performance improvement. To evaluate the improvement from adding a forecaster, OW are learned with the selected forecasters and tested in a cross-validation procedure. Traditionally, forward feature selection uses a stopping condition that adds more features until no improvement in model performance can be measured. To remain consistent with the previously proposed information criteria, the use of a stopping condition is omitted and forecasters are added until all forecasters are selected. However, the criterion could also be adjusted so that a certain performance threshold must be reached when adding a forecaster, thereby pruning the candidate selection vectors and reducing the search space for tuning the model. Compared to the benchmark models OW, SA, models for linear and non-linear shrinkage from OW to SA, and the alternative integrated model peLASSO, NLHS shows the largest improvements in terms of MSE on unseen data with an increasing number of forecasters of twelve or more, averaged over all other simulation parameters such as correlation, training size, and variances. The second best-performing model in this setting represents the non-linear shrinkage from OW to SA. The results of the study thus confirm the motivation of the work. With a high number of forecasters, the model instability of OW increases, and the probability of learning one or more extremely incorrect OW grows. The flexible regularization path of the non-linear shrinkage models provides the most effective approach to correct the erroneous OW, where the NLHS models with performance-based selection and forward feature selection lead to further accuracy improvements by shrinking OW to SubA instead of solely shrinking OW to SA.

3. Research Contribution

The third paper proposes an extended version of the forward feature selection criterion presented in Contribution 2 to account for diversity gains in the selection process. The motivation for this paper is based on the relationship between the degree of diversity among forecasters and the performance of forecast combinations (Atiya, 2020; Lichtendahl & Winkler, 2020; Kang et al., 2022). An important reason for this relationship is that the selection of diverse forecasters, among others, increases the likelihood of suppressor effects occuring that positively affect the predictive power of other forecasters. As in the previous forward feature selection procedure, the proposed criterion starts with the selection of the best forecaster in the sample. In the new procedure, however, not only the accuracy gain, but also the gain in diversity added by the inclusion of a candidate in the selection is being measured. The accuracy gain is calculated as the ratio between the MSE after adding a candidate and the previously obtained MSE, which is calculated based on OW combinations in cross-validation. Diversity gain, on the other hand, is measured by the root of the multiple coefficient of determination obtained by regression on the error data with the candidate as the dependent variable and the previously selected candidates as independent variables. Accuracy gain and diversity gain are combined multiplicatively, with a hyperparameter allowing to prioritize diversity over accuracy.

Error data from eight forecasters are generated for the simulation. As the only one of the four papers, the correlation values between the forecasters are not modeled identically in pairs, but are drawn from a correlation range and assigned to the forecasters in such a way that forecasters with worse performance are more diverse to other forecasters. Thereby, the design of the simulation is oriented on the findings of Kourentzes et al. (2019), which emphasize that mainly accurate forecasters should be selected for the combination, although comparatively poor forecasters can also be included if they introduce more diversity into the pool of selected forecasters. In addition, all previously introduced selection criteria are applied.

Simulation results show that the proposed criterion performs particularly well in combination with LHS when the number of training data and variance differences between forecasters are high. Whereby, the results further show that in areas of high correlation, the setting of the hyperparameter in the new selection criterion has no effect on the results. On the other hand, for low to medium correlations, slightly better results are obtained in the MSE on test data compared to the other criteria when diversity is weighted more heavily than accuracy.

4. Research Contribution

The fourth paper is the main contribution of this work and proposes a one-stage model for hybrid shrinkage of OW toward SubA. Although the presented models LHS and NLHS show promising results in extensive simulations, the motivation of the paper stems from the dependence of the performance of the two-stage models on the preceding forecaster selection step.

As the number of forecasters grows, the number of possible combinations of subset average weights increases exponentially. To identify the subset average weights which provide the best shrinkage targets for regularization, the proposed information criteria offer a reasonable approach for a priori pruning the search space. However, the use of different information criteria may lead to varying selection results in comparable forecast settings, which adds complexity to the practical implementation of LHS and NLHS and raises questions about the optimality of the two-stage approach.

A solution is presented by the model *Hybrid Shrink* (HS), which allows the selection and weighting of forecasts to be performed simultaneously rather than sequentially. The model uses regularized regression that combines an OLS term with an L_1-penalty and $L_{1/\beta}$-penalty term. The L_1-penalty term penalizes absolute weight deviations from a given (subset) mean, while the $L_{1/\beta}$-penalty term with a sufficiently small $\beta > 1$ penalizes absolute weight deviations from zero. With the function learning itself which forecaster weights to be regularized toward equality, HS non-linearly shrinks the OW of forecasters toward SubA. To control the number and degree of shrinkage of forecasters to the mean, the model requires the tuning of two hyperparameters.

The flexibility of the function to determine the shrinkage direction of the forecasters by itself, though, complicates the hyperparameter tuning via a classical cross-validation process. Within different foldings of cross-validation, different forecasters may be shrunk toward the mean of the subset, which distorts the evaluation of shrinkage of OW toward SubA on the entire dataset by classical cross-validation. This observation is amplified by a phenomenon we refer to as "over-shrinkage". Over-shrinkage describes that regularization models tend to shrink too much when data availability is limited as small batches of validation data in the cross-validation process cannot reflect the true underlying error structure.

As a solution, an adapted cross-validation procedure is introduced for HS, which first learns the SubA combination with maximum shrinkage on the whole data and fixes it for the cross-validation process. The aim of this new procedure is to stabilize the learning process of the shrinkage paths within cross-validation and to allow a more reliable determination of the required shrinkage.

Based on a large-scale experimental study with 320 forecast scenarios, the presented HS method with its new sampling procedure is analyzed, where each scenario is repeated 25 times. OW, SA, SubA, linear and non-linear shrinkage via eLASSO from OW to SA are used as benchmark methods. For the analysis of the results, a decision tree is again learned based on the simulation parameters and the average best model across all replications with respect to the MSE out-of-sample as the target variable. In 135 out of 320 scenarios, HS achieved the lowest average MSE on the test data, representing 42 % of all scenarios. Among these, similar to previous hybrid shrinkage methods, HS shows the best results in prediction scenarios with a large number of 10, 12, or 15 forecasters. Here, HS can reduce the MSE test value of OW by up to 68 % with a small number of training observations. Further investigation of the shrinkage behavior of HS confirms that the new sampling procedure produces lower levels of shrinkage compared to a conventional cross-validation procedure. Thereby, the lower shrinkage leads to an improvement of the MSE on unknown data, especially when the number of training data is small. On the other hand, when the data availability improves, the shrinkage ratios of HS with classical cross-validation and the proposed procedure become increasingly similar. This suggests that higher data availability facilitates learning the same shrinkage targets within the cross-validation process, but simultaneously highlights the relevance of the new sampling procedure for more reliable estimation of the required shrinkage in scenarios with high uncertainty.

3 Publication Details

The cumulative dissertation consists of four individual scientific papers, with the first three papers representing peer-reviewed conference papers and the last a submitted journal paper. All four contributions are in accordance with the doctoral regulations of the Faculty of Business and Economics of the Katholische Universität Eichstätt-Ingolstadt as amended on February 14, 2012, and their amendments in the bylaws of April 30, 2015, December 20, 2017, and September 13, 2019. The following chapter presents publication details, which include information on authorship, publication, and conference presentation.

Contribution 1: Linear Hybrid Shrinkage of Weights for Forecast Selection and Combination

- Authorship: This paper is joint-work at the Katholische Universität Eichstätt-Ingolstadt of Felix Schulz, Thomas Setzer and Nathalie Balla with editing shares as shown in Table 1.
- Publication: The paper was published online on January 4, 2022, in the Proceedings of the 55th Hawaii International Conference on System Sciences [Ranking: VHB-JQ3: C, CORE2018: A], doi: 10.24251/HICSS.2022.267.
- Conference Presentation: After the paper was accepted, it was presented and discussed by the first author at the 55th Hawaii International Conference on System Sciences which took place from January 4-7, 2022. Due to the COVID-19, the presentation was held digitally.

Table 1: Editing shares of the authors of Contribution 1

Task	F. Schulz	T. Setzer	N. Balla
Literature review	60 %	20 %	20 %
Model development	60 %	20 %	20 %
Research design	60 %	20 %	20 %
Implementation	60 %	20 %	20 %
Data analysis	60 %	20 %	20 %
Result interpretation	60 %	20 %	20 %
Writing process	60 %	20 %	20 %
Total	**60 %**	**20 %**	**20 %**

Contribution 2: Non-Linear Hybrid Shrinkage of Weights for Forecast Selection and Combination

- Authorship: This paper is single-authored by Felix Schulz.
- Publication: The paper was published online in the Proceedings Wirtschaftsinformatik 2022 on January 17, 2022 [Ranking: VHB-JQ3: C, CORE2018: C], URL: `https://aisel.aisnet.org/wi2022/business_analytics/business_analytics/7`.
- Conference Presentation: After acceptance, the paper was presented and discussed at the 17th International Conference on Wirtschaftsinformatik, hosted digitally on February 21-23, 2022, due to COVID-19. The paper was awarded 3rd place in the category 'Best Short Paper'.

Contribution 3: A Combined Measure Based on Diversification and Accuracy Gains for Forecast Selection in Forecast Combination

- Authorship: This paper is joint-work at the Katholische Universität Eichstätt-Ingolstadt of Felix Schulz, Thomas Setzer and Nathalie Balla with editing shares as shown in Table 2.
- Publication: The final version of the paper has not yet been published, but was accepted for the Operations Research Proceedings on September 30, 2022 [Ranking: VHB-JQ3: D, CORE2018: C].
- Conference Presentation: A presentation was held by the first author at the International Conference on Operations Research (OR2022) in Karlsruhe, Germany, September 6-9, 2022.

Table 2: Editing shares of the authors of Contribution 3

Task	F. Schulz	T. Setzer	N. Balla
Literature review	70 %	20 %	10 %
Model development	70 %	20 %	10 %
Research design	70 %	20 %	10 %
Implementation	70 %	20 %	10 %
Data analysis	70 %	20 %	10 %
Result interpretation	70 %	20 %	10 %
Writing process	70 %	20 %	10 %
Total	**70 %**	**20 %**	**10 %**

Contribution 4: Shrinkage of Weights Towards Subset Selection in Forecast Combination

- Authorship: This paper is joint-work at the Katholische Universität Eichstätt-Ingolstadt of Felix Schulz and Thomas Setzer with the editing shares as shown in Table 3.
- Publication: The paper was submitted to the Journal of Business Research [VHB-JQ3: B] and is in the review process.
- Conference Presentation: /

Table 3: Editing shares of the authors of Contribution 4

Task	F. Schulz	T. Setzer
Literature review	80 %	20 %
Model development	70 %	30 %
Research design	70 %	30 %
Implementation	70 %	30 %
Data analysis	70 %	30 %
Result interpretation	70 %	30 %
Writing process	60 %	40 %
Total	**70 %**	**30 %**

4 Abstracts of Paper

The following chapter presents the abstracts of the four papers submitted for the cumulative dissertation. To improve readability, the abstracts are presented on separate pages.

Paper 1

Linear Hybrid Shrinkage of Weights for Forecast Selection and Combination

Felix Schulz, Thomas Setzer, Nathalie Balla

Abstract: Forecast combination is an established methodology to improve forecast accuracy. The primary questions in the current literature are how many and which forecasts to include (selection) and how to weight the selected forecasts (weighting). Although integrating both tasks seems appealing, we are only aware of a few data analytical models that integrate both tasks. We introduce Linear Hybrid Shrinkage (LHS), a novel method that uses information criteria from statistical learning theory to select forecasters and then shrinks the selection from their in-sample optimal weights linearly towards equality, while shrinking the non-selected forecasts towards zero. Simulation results show conditions (scenarios) where LHS leads to higher accuracy than LASSO-based Shrinkage, Linear Shrinkage of in-sample optimal weights, and a simple averaging of forecasts.

Keywords: Forecast combination, forecast selection, lasso, shrinkage, variable importance

Published: Proceedings of the 55th Hawaii International Conference on System Sciences
URL: `https://hdl.handle.net/10125/79599`

Paper 2

Non-Linear Hybrid Shrinkage of Weights for Forecast Selection and Combination

Felix Schulz

Abstract: We introduce Non-Linear Hybrid Shrinkage (NLHS) as a holistic model for forecast combination, shrinkage and selection. NLHS first determines the selection of forecasters based on information criteria such as forward feature selection and stores the selection status of forecasters in a selection vector. Depending on the selection status, the estimated optimal weights of the forecasters are either shrunk to zero or equal weights by the least absolute shrinkage and selection operator (LASSO). Among benchmark algorithms such as simple average, optimal weights, or linear and LASSO-based shrinkage models, NLHS is superior for a larger number of forecasters, as shown in simulation-based experiments.

Keywords: Forecast combination, shrinkage, forward feature selection

Published: Wirtschaftsinformatik 2022 Proceedings. 7.
URL: `https://aisel.aisnet.org/wi2022/business_analytics/business_analytics/7`

Paper 3

A Combined Measure Based on Diversification and Accuracy Gains for Forecast Selection in Forecast Combination

Felix Schulz, Thomas Setzer, Nathalie Balla

Abstract: Recent innovations in the field of forecast combination include integrated models for forecast selection, weighting and regularization. The models proposed in related articles first label whether or not forecasters should remain in the selection using information criteria from statistical learning theory. Depending on the selection status, the optimal weights of all forecasters in the sample are then used as baseline to shrink the weights either toward zero or the mean, with the degree of regularization determining the final selection of forecasters. In this paper, we propose a new information criterion reflecting the importance of diversification and accuracy gains in the selection of forecasters for integrated models. In an iterative procedure motivated by forward feature selection, each forecaster is selected sequentially, while at each step the increase in accuracy and diversification due to the addition of a forecaster to the previous selection is measured. To quantify the increase in diversity, the multiple correlation coefficient is used, which captures the correlation between the previously selected forecasters and a candidate, where the lower the correlation between the candidate and the selection, the higher the gain in diversity for the combination. For the accuracy increase, the accuracy achieved by optimal weight combinations with the previously selected forecasters is compared with the accuracy after adding a candidate. A hyperparameter further enables the trade-off between accuracy and diversification gains in the criterion. Simulation-based studies show scenarios in which our presented information criterion achieves advantages in out-of-sample prediction accuracy over previous criteria for selection by accounting for accuracy and diversification gains.

Keywords: Forecast algorithms, Accuracy–Diversity, Forecast combination

Accepted: Operations Research Proceedings 2022
URL: `NA`

Paper 4

Shrinkage of Weights Towards Subset Selection in Forecast Combination

Felix Schulz, Thomas Setzer

Abstract: We propose Hybrid Shrink (HS), a novel model to combine forecasts that merges the concepts of forecast selection, weighting, and shrinkage by shrinking forecaster weights either to their average weight, i.e., towards subset selection, or to zero. HS learns which forecasts to shrink in which direction by minimizing the Mean Squared Error (MSE) of the combined forecast on past error data plus two penalty terms. Two hyperparameters are used to control the proportion of MSE versus penalty costs and the number of forecasts to be shrunk to their subset average. We further propose a novel procedure to tune the hyperparameters that learns shrinkage paths using all available training data, while the shrinkage level is then determined subsequently using cross-validation. The performance of HS is evaluated on a comprehensive set of simulated forecast error data. Results demonstrate the capability of HS to reduce MSE compared to approaches like taking the simple average, performing subset-averaging, or using shrinkage towards average weights and standard cross-validation to learn and shrink weights.

Keywords: Forecast combination, Weight shrinkage, Subset selection, Bias–Variance tradeoff

Submitted: Journal of Business Research
URL: NA

5 Conclusion

Forecast combination is a widely used strategy to increase the accuracy of individual models for out-of-sample forecasts. This paper presents a family of integrated models for combining the tasks of forecast selection, weighting, and shrinkage, and extends the existing literature on forecast combination in a meaningful way. Compared to simple regularization methods that shrink OW toward SA, the proposed hybrid shrinkage models provide great flexibility to adjust incorrect or overfitted optimal weights by shrinking not only toward the mean but also toward zero. Through large-scale simulations, the work provides considerable evidence that shrinkage towards subset average weights can lead to asymptotic improvements in out-of-sample MSE. Across all simulations, a clear picture emerges for the practical implementation and theoretical considerations of integrated methods with the highest improvements in forecast accuracy in scenarios with many forecasters relative to the available training data.

References

Aiolfi, A., & Timmermann, A. (2006). Persistence in Forecasting Performance and Conditional Combination Strategies. *Journal of Econometrics*, *135*, 31 – 53.

Atiya, A. F. (2020). Why does forecast combination work so well? *International Journal of Forecasting*, *36*(1), 197 – 200.

Banerjee, A., Marcellino, M., & Masten, I. (2005). Forecasting macroeconomic variables for the new member states of the European Union. *European Central Bank. Working paper Series*, *482*.

Bates, J., & Granger, C. (1969). The Combination of Forecasts. *Journal of the Operational Research Society*, *20*(4), 451 – 468.

Blanc, S., & Setzer, T. (2020). Bias–Variance Trade-Off and Shrinkage of Weights in Forecast Combination. *Management Science*, *66*(12), 5485 – 6064.

Bojer, C. S., & Meldgaard, J. P. (2021). Kaggle forecasting competitions: An overlooked learning opportunity. *International Journal of Forecasting*, *37*(2), 587 – 603.

Diebold, F., & Pauly, P. (1990). The use of prior information in forecast combination. *International Journal of Forecasting*, *6*(4), 503 – 508.

Diebold, F., & Shin, M. (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives. *International Journal of Forecasting*, *35*(4), 1679 – 1691.

Geweke, J., & Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, *164*(1), 130 – 141.

Granger, C. W. J. (2014). *Forecasting in business and economics*. Academic Press.

Kang, Y., Cao, W., Petropoulos, F., & Li, F. (2022). Forecast with forecasts: Diversity matters. *European Journal of Operational Research*, *301*(1), 180 – 190.

Kourentzes, N., Barrow, D., & Petropoulos, F. (2019). Another look at forecast selection and combination: Evidence from forecast pooling. *International Journal of Production Economics*, *209*, 226 – 235.

Lichtendahl, K., & Winkler, R. (2020). Why do some combinations perform better than others? *International Journal of Forecasting*, *36*(1), 142 – 149.

Makridakis, S., Fry, C., Petropoulos, F., & Spiliotis, E. (2022). The future of forecasting competitions: Design attributes and principles. *INFORMS Journal on Data Science*, *1*(1), 96 – 113.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, *34*(4), 802 – 808.

Makridakis, S., & Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science*, *29*(9), 987 – 996.

Mannes, A., Soll, J., & Larrick, R. (2014). The Wisdom of Select Crowds. *Journal of Personality and Social Psychology,*, *107*(2), 276 – 299.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, *23*(2), 821 – 862.

Smith, J., & Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, *71*(3), 331 – 355.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301 – 320.