Combining professional forecasts: A closer look at partially-egalitarian Lasso

Justin Dijk - 457469
Supervisor: P.A. Opschoor
Second Assessor: Dr. A. Tetereva

July 4, 2020

Abstract

This research further explores the use of partially-egalitarian Lasso for combining professional forecasts obtained through surveys. Because expert forecasts are highly correlated, we argue that utilizing Lasso as a trimming method in the first step of this technique is not optimal. In an ex-post analysis with various alternative selection methods, we find that trimming with the Elastic Net outperforms trimming with Lasso. However, even with the Elastic Net, we confirm previous findings that discarding most of the forecasts and averaging the remainder is optimal. Additionally, we find that the selected subset of forecasts is highly dependent over time. Based on these results, we introduce an ex-ante method that tunes the penalty parameters with a hold-one-out approach. This ex-ante procedure performs as well or better than the previously successful subset-averaging approach but is computationally much cheaper. As a consequence, it remains of practical use even with a large number of forecasters.

Bachelor Thesis Econometrics and Operations Research
Erasmus University Rotterdam
Erasmus School of Economics

Contents

1 Introduction

Since the seminal paper by Bates and Granger (1969), it is well known that combinations of forecasts routinely outperform individual forecasts. There are various possible reasons for this success[1]. One particular reason is that forecasters can have different underlying information sets or modeling approaches. This means that by combining different predictions, information sets are pooled and potential diversification, against model-specific biases and unknown misspecifications, takes place.

Although the idea of forecast combinations has a long history, how to combine them is still much disputed. Empirically, simply taking an average of the forecasts often outperforms more sophisticated and theoretically optimal weighting schemes. A finding that mostly gets attributed to large estimation errors of the combining weights (Clemen and Winkler (1986); Hendry and Clements (2004)).

Another unresolved issue is whether each forecast should be included or not. As shown in Winkler and Makridakis (1983)), it is often the case that it is beneficial to at least exclude very poor performing predictors, especially in an equal-weighted combination. There have also been successful results with taking a more aggressive approach of trimming up to 80 percent, thus only including the top 20 percent performers in the combination (Aiolfi and Favero (2005)).

Given these findings, Diebold and Shin (2019) introduce a one-step Lasso-based approach, in which some candidate forecasters are excluded and the remaining combination weights are shrunk towards equality. This approach, aptly named the partially-egalitarian Lasso (peLasso), is however difficult to solve in one step, hence the implementation is done in two phases. First, use a selection method to decide which forecasts to give a non-zero weight. Then, shrink these weights towards equality with, for example, a rescaled Lasso or Ridge procedure.

They apply peLasso to GDP growth forecasts of the European Central Bank Survey of Professional Forecasts and find that, in an ex-post analysis, most predictions should be trimmed and the remainder averaged. In light of these results, they propose an ex-ante subset-averaging approach that does not require the choice of a tuning parameter. This ex-ante procedure turns out to perform significantly better than taking a simple average over all of the candidate forecasts and can be used in real-time forecasting.

There are, however, a few potential shortcomings with their approach and results. First off, they decide to use Lasso as a selection method to reduce the number of forecasters to combine. Even though Lasso performs generally better than a stepwise selection method, it often breaks down when candidate forecasts are highly correlated (Zhao and Yu (2006)). This is a potential problem because high correlations are a likely situation when forecasts come from a large number of models or professional surveys, a point also made by Roccazzella et al. (2020). Secondly, as already mentioned in Diebold and Shin (2019), the comparative performance of their peLasso method strongly relies on the dataset that they use. This dataset is indirectly also of vital importance for their motivation behind their subset-averaging approach because it relies on the result, of the ex-post analysis, that remaining forecasts should be averaged. Consequently, it remains to be seen whether this ex-ante

[1]For a summary, see Timmermann (2006) and Elliott and Timmermann (2008).

approach performs well in different forecast survey datasets. Lastly, the practical applicability of the subset-averaging method is questionable because it involves calculating the past performance of every possible subset of forecasters, which can quickly become computationally too expensive when there are many forecasters.

In this research, we address these shortcomings. Specifically, we challenge whether selection methods that can deal with high correlations, improve or change the main results found in the ex-post analysis of Diebold and Shin (2019), namely that heavy trimming and averaging is optimal for forecast survey data. This leads to the second question of how we can construct, with these new results, a real-time forecasting method that is computationally less expensive and still provides better results than taking a simple average.

Consequently, with these questions, we extend the currently existing literature in two meaningful ways. Firstly, we further address the trimming and averaging of forecasts and how to do so in the case that forecasts are highly correlated. This is not only valuable for forecasts of professional surveys but also for when a large number of forecasting models are considered such that high correlations are likely to be present. Secondly, by constructing an ex-ante method that is computationally less expensive, we provide a new way of combining forecasts that remains useful in practice.

In our ex-post analysis, we find that trimming with the Elastic Net provides better results than trimming with Lasso across multiple professional survey datasets. However, preconditioning the forecasts with the (Generalized) Puffer transformation to make them less correlated, is not performance-enhancing. More importantly, with the Elastic Net, the main findings in Diebold and Shin (2019) still hold, namely that harsh trimming and averaging the remaining forecast is optimal. With these findings and the fact that the selected subset of forecasters in the optimal solution is dependent over time, we introduce an ex-ante tuning method that uses a hold-one-out principle. This method provides similar or better results for all datasets than subset-averaging, it is a more straightforward implementation of partially-egalitarian Lasso and is computationally much cheaper, which means that it remains applicable in practice even when datasets have a large number of forecasters.

The layout of the remainder of this text is as follows. In Section 2, we discuss partially-egalitarian Lasso and the problems with Lasso as a selection method. Additionally, in this section, we introduce two alternatives to Lasso and address how to tune the parameters both ex-post and ex-ante. Next, in section 3, we present the survey data for the empirical applications. Section 4 applies the different trimming methods, in which the tuning parameters are optimized ex-post and takes a closer look at the set of selected forecasters in the optimal solution. In section 5, we show the results for the ex-ante tuning methods. Lastly, section 6 concludes.

2 Methodology

2.1 The Forecast Combination Problem

Suppose we want to make a point forecast at time T for a variable of interest one period ahead with realization y_{T+1}. Additionally, we have multiple models at our disposal that provide forecasts $x^1_{T+1}, x^2_{T+1}, \cdots, x^P_{T+1}$, where x^i_{T+1} is the prediction of model i for y_{T+1}. The forecast combination problem comes down to creating one single point forecast $\hat{y}_{T+1}$ out of the P predictions, such that a loss function gets minimized.

This loss function is generally assumed to be increasing in the forecast error e_{T+1}. However, it can have vastly different shapes due to the preferences of the forecaster. For example, he or she may find overestimating more costly than underestimating the variable of interest, such that the function is asymmetric as opposed to symmetric. Because the loss function determines for a large part how to optimally combine forecasts (Timmermann, 2006), we assume that it is simply equal to the Mean Squared Error (MSE).

Furthermore, we only consider linear combinations of the predictions, which means that the forecast combination problem boils down to finding the optimal weights $\beta_1, \cdots, \beta_P$ in the equation

$$\hat{y}_{T+1} = \beta_1 x^1_{T+1} + \beta_2 x^2_{T+1} + \cdots + \beta_P x^P_{T+1}. \tag{1}$$

Under the above assumptions the combining weights are often estimated by a simple Ordinary Least Squares (OLS) regression of y_{T+1} on $x_{T+1} = (x^1_{T+1}, \cdots, x^P_{T+1})$ with realized data of the previous periods. When the individual predictors are presumed to be unbiased one could further add a constraint that the weights need to sum to one. In fact, Granger and Ramanathan (1984) show that in this case the theoretically optimal weights can be recovered through a simple OLS estimation.

2.2 Partially-egalitarian Lasso

As was mentioned before, in practice, estimating optimal weights often has a worse performance than simply taking an average of all the forecasts. Additionally, it is not clear whether each forecast should be included in the combination because some may not be much more than noise.

For these reasons Diebold and Shin (2019) introduce the partially-egalitarian Lasso (peLasso). Let $y = (y_1, \cdots, y_T)$, $\beta = (\beta_1, \cdots, \beta_P)$ and X be the data matrix consisting of the forecast vectors $x_1, \cdots, x_T$. Then the peLasso estimator is defined as

$$\beta_{\text{peLasso}} = \arg\min_{\beta} \Big(\|y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \sum_{i=1}^{P} |\beta_i - \frac{1}{p(\beta)}| \Big), \tag{2}$$

where $\|\cdot\|_q$ is the L_q-norm and $\lambda_1, \lambda_2 \geq 0$.

The second term, in the above penalized linear regression, is the well known L_1-norm penalty; as the tuning parameter λ_1 increases, the combination weights get increasingly shrunk towards zero and some weights can become equal to zero. In other words, the first penalty provides the trimming of some forecasters, and the amount of trimming that is done increases when the value of the tuning parameter λ_1 increases. Also note that the first two terms together form the Lasso estimator, a much-used selection method (Tibshirani, 1996).

Because averaging performs really well in practice, the L_1 penalty shrinks in the wrong direction, namely to zero. That is why Diebold and Shin (2019) add a second penalty term to Lasso, in which $p(\beta)$ is the number of non-zero weights. Consequently, this second penalty shrinks the remaining non-zero weights towards an average and this shrinkage becomes heavier as the tuning parameter λ_2 becomes larger.

The above minimization is hard to solve due to the discontinuities at $\beta_i = 0$. However, the solution can be found with a two-step approach instead. In the first step, use a selection method to select a subset of p forecasters out of P candidate forecasters. The second step is to use a standard method on the p surviving forecasts, to estimate their combination weights and shrink them towards $1/p$.

Diebold and Shin (2019) utilize Lasso to select the forecasters and eRidge, eLasso or an average for the estimation of the combining weights in the second step[2]. The decision of using Lasso as the selection method may not be optimal, which we turn to next.

2.3 Lasso Selection Inconsistency

It has been pointed out in several papers that the Lasso L_1-norm regularization is only selection consistent when a rather stringent condition holds on the correlation between the predictors (Meinshausen et al., 2006; Zou, 2006; Zhao and Yu, 2006). The following definition of selection consistency is used in these papers.

Definition 1. An estimator is **selection consistent** if it holds that

$$P\big(\{i : \hat{\beta}_i^n \neq 0\} = \{i : \beta_i^n \neq 0\}\big) \rightarrow 1 \text{ as } n \rightarrow \infty. \tag{3}$$

In other words, when the sample size approaches infinity, the variables with an estimated nonzero weight equals the set of variables with nonzero weights under the true model. Note that, this is different from parameter estimation consistency where the actual values of the estimated coefficients equal the true values. With regards to peLasso, selection consistency is appropriate because the optimal coefficients are separately determined in step two after the selection has taken place.

Zhao and Yu (2006) found a necessary condition on the design matrix X for Lasso to be selection consistent, called the Irrepresentable Condition.

Definition 2. Assume that the matrix X can be subdivided into two parts X_S and X_{S^c}, where X_S corresponds with the covariates of the true support and X_{S^c} its complement. The design matrix X

[2] eRidge and eLasso are similar to standard Ridge and Lasso methods but are scaled to shrink towards $1/p$ instead of 0, see Diebold and Shin (2019) for a full discussion of these shrinking methods.

satisfies the **Irrepresentable Condition** if there exists a constant $\eta \in (0,1]$ such that

$$\left\| X_{S^C}^T X_S (X_S^T X_S)^{-1} \text{sign}(\beta_s) \right\|_\infty \leq 1 - \eta. \tag{4}$$

In the above equation, $\| \cdot \|_\infty$ is the max norm and $\text{sign}(\beta_s)$ the sign of the coefficients in the support of the true model.

Although this definition is quite technical, it resembles a restraint on the regression coefficients of the irrelevant predictors with the relevant predictors, namely that these cannot be too large. In reality, the true support is not known, which means this condition is difficult to check. However, a data matrix that contains highly correlated predictors is more likely to fail the above condition because, due to the high correlations, it is more likely that the regression coefficients between relevant and irrelevant variables are large. If this is indeed the case, Lasso tends to simply select one variable out of a group of correlated variables and does this rather randomly, which can lead to poor forecasting performances (Zou and Hastie, 2005; Zhang et al., 2010). Considering that professional forecasts are often found to be highly correlated, it is meaningful to see if selection methods that can deal with this correlation, provide different or perhaps better results than found in Diebold and Shin (2019).

2.4 Alternatives to First-step Lasso

Besides being able to deal with high correlations, there are two additional requirements for the trimming methods in this paper. In most professional surveys, a large number of professional institutions are asked for their forecasts. Additionally, it is oftentimes beneficial to use small (moving) estimation windows to determine the optimal combination weights (Winkler and Makridakis, 1983; Bates and Granger, 1969; Newbold and Granger, 1974). Hence the first additional requirement is that the trimming method can deal with situations where $T > P$. This situation also occurs in the empirical applications in sections 4 and 5. The second requirement is that the selection method does not directly use the variance-covariance matrix of the predictors as these methods tend to perform poorly in small sample sizes (Mundfrom et al., 2005).

2.4.1 Elastic Net

The first alternative to Lasso is the Elastic Net (EN). A method that was introduced by Zou and Hastie (2005) to address two downsides of Lasso its variable selection capabilities.

(a) When $P > T$, Lasso can select at most T variables.

(b) If there are variables with high pairwise correlations, Lasso tends to select only one of them and does so rather arbitrarily.

Note that when $P > T$, the first limitation rules out that peLasso finds a solution in which a simple average with no trimming is the best weighting scheme. This can be a problem because there are theoretical situations in which taking a simple average over all forecasters is optimal (Elliott, 2011).

The Elastic Net estimator is defined as follows

$$\beta_{\text{EN}} = \arg\min_{\beta}\Big(\|y - X\beta\|_2^2 + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2\Big). \tag{5}$$

Above criterion has two different penalty terms with tuning parameters λ_1 and λ_2.

The first penalty is the L_1-norm penalty that is also used in Lasso, which means that the Elastic Net is still capable of selecting variables by putting certain coefficients equal to zero. The second one is the L_2 norm, which is also found in Ridge regression. This penalty is not capable of selecting variables but only shrinks them towards zero and is found to perform well under multicollinearity (Hoerl and Kennard, 1970).

In summary, the L_1-norm provides sparsity while the L_2-norm provides stability under multicollinearity. For these reasons, the Elastic Net is a good alternative to Lasso when there are strong pairwise correlations between the predictors. A downside, however, is that it introduces another tuning parameter that has to be optimized.

2.4.2 Adaptive Elastic Net

Zou and Zhang (2009) point out that the Elastic Net does not possess the oracle property[3]. Subsequently, they introduce the Adaptive Elastic Net (AEN) estimator, which weights the terms in the L_1 penalty.

Suppose that first the regular Elastic Net is performed, which gives the estimated coefficients $\hat{\beta}_i$ for $i = 1, \cdots, P$. Then the Adaptive Elastic Net estimator is defined as

$$\beta_{\text{AEN}} = \arg\min_{\beta}\Big(\|y - X\beta\|_2^2 + \lambda_1^* \sum_{i=1}^{P} \hat{w}_i\beta_i + \lambda_2\|\boldsymbol{\beta}\|_2^2\Big), \tag{6}$$

where $\hat{w}_i = (|\hat{\beta}_i|)^{-\gamma}$ for a positive constant γ. To avoid dividing by zeros $\hat{w}_i = \infty$ when $\hat{\beta}_i = 0$. Note that the tuning parameter λ_1^* in above equation is allowed to be different from λ_1 in the estimation of the initial Elastic Net. However, this is not the case for the L_2-norm penalty parameter λ_2.

The Adaptive Elastic Net retains the properties of the regular Elastic Net method but tends to set smaller coefficients equal to zero because of their corresponding large weights in (6). In practice, this could mean that it trims more and thus chooses a different set of optimal forecasters. That is why this version is also of importance as an alternative for Lasso. Note also, that the total amount of tuning parameters now equals four.

2.4.3 The Puffer Transformation

Another approach is to stick with Lasso but transform the data in such a way that the irrepresentable condition is likely to hold. A clear benefit of this preconditioning, compared to the (Adaptive) Elastic Net, is that no extra tuning parameters have to be optimized.

[3]For a full discussion on the oracle property and its benefits see Leeb and Pötscher (2008).

In most cases, the data is transformed by left multiplying the equation of interest by an appropriate matrix F such that the transformed linear equation becomes

$$Fy = (FX)\beta + F\epsilon, \tag{7}$$

where ϵ is a vector of error terms.

There is a strong similarity between this approach and Generalized Least Squares, where the matrix F is chosen in such a way that the above error terms are homoskedastic and uncorrelated. However, in this case, the aim is not to adjust the error terms, but to make FX satisfy the irrepresentable condition such that Lasso is model selection consistent. Importantly, this transformation also does not change the vector of unknown weights β.

Jia et al. (2015) suggest to use the Puffer transformation $F = UD^{-1}U^T$, in which D and U are matrices from the Singular Value Decomposition of $X = UDV^T$. As a consequence, D is a diagonal matrix with the singular values of X, and U and V are orthogonal matrices that consist of its left and right singular vectors, respectively. They show through simulation that the transformed data matrix FX has drastically reduced pairwise correlations, meaning it is more likely to comply with the irrepresentable condition[4].

2.4.4 Generalized Puffer Transformation

A potential downside of the Puffer transformation is that it can blow up the variance of the error term. For example, if $\epsilon \sim N(0, \sigma^2 I_T)$, then the covariance matrix of $F\epsilon$ is equal to $\sigma^2 UD^{-2}U^T$, which means that the noise is strongly amplified if a singular value of X is close to zero. Put differently, there is a trade-off between a well-behaved data matrix and a well-behaved error term. If this trade-off is too large, the benefit of performing Lasso on the transformed data (FX, Fy) is nullified.

To prevent the variance of becoming too large Jia et al. (2015) also introduce the Generalized Puffer (GPuffer) transformation which uses the matrix $F = U\tilde{D}U^T$, where $\tilde{D} = f(D_{ii}, \tau)/D_{ii}$. In this case f is a hard thresholding function

$$f(D_{ii}, \tau) = 1 \text{ if } D_{ii} > \tau \text{ and zero otherwise.} \tag{8}$$

The idea behind this generalized preconditioner is to set an upper bound on the diagonal entries in the matrix D^{-1}, namely τ^{-1}. As a consequence, the amplification of the disturbance terms is also bounded. Importantly, the parameter τ balances the trade-off between the size of the error terms and the likeliness that the transformed design matrix satisfies the irrepresentable condition. Jia et al. (2015) provide a theorem that τ needs to be greater than or equal to $\sqrt{\frac{P}{T}}$ for this likeliness to become large enough to defend the use of such a hard thresholding function. Because of this, we only consider these possible values for the thresholding parameter in our empirical applications.

[4] In a low dimensional setting $P < T$ and with singular values bounded away from zero the irrepresentable condition is always satisfied (Jia et al., 2015).

2.5 Parameter Tuning

To be able to perform the various peLasso methods we need to determine, for each method separately, the value of the penalty parameter λ. Of course in the case that we have multiple tuning parameters, λ is a vector. We discuss three different approaches; Ex-post tuning, subset-averaging and hold-one-out tuning.

2.5.1 Ex-post Tuning

In ex-post tuning, we choose the value of λ that minimizes the out of sample Root Mean Squared Error (RMSE), which we can discover ex-post by a simple grid search. Hence, we give ourselves valuable information that we would not have if we were to combine forecasts in real-time. Additionally, we only choose one optimal value of λ that minimizes the RMSE for the whole sample, instead of optimizing it for each estimation window separately. Diebold and Shin (2019) use this type of tuning to motivate their subset-averaging method that is completely ex-ante. We apply ex-post tuning in section 4.

2.5.2 Subset-averaging

Given the results of the ex-post analysis for their GDP growth forecasts dataset, Diebold and Shin (2019) introduce a method that does not require any tuning of parameters and can be implemented in real-time. The motivation behind this process fully relies on the result that harsh trimming and averaging of the surviving forecasts is optimal.

This ex-ante subset-averaging process can be described as follows. At each period, rolling forward, select the subset of forecasters, with a maximum size of $N_{\max}$, of which its average has performed the best over the estimation window with length W. The prediction for that period is then simply the subset-average of that best performing subset. In other words, use that particular average that has performed the best over the previous periods. It is immediately clear why this method can become computationally too expensive to be of any practical use when $N_{\max}$ or the total amount of forecasters grows large. We perform subset-averaging in subsection 5.1.

2.5.3 Hold-one-out Tuning

Perhaps surprisingly, Diebold and Shin (2019) forego the option of tuning the penalty parameters on the estimation window. Doing so would give a more direct way of implementing peLasso ex-ante. There are many popular ways to tune the penalty parameters for Lasso or Elastic Net, like k-fold or leave-one-out cross-validation. However, most of these popular tuning methods ignore the temporal dependencies of the dataset (Tashman, 2000). This is problematic because Diebold and Shin (2019) show in their ex-post results that the selected subset of forecasters is highly dependent over time.

To overcome this problem we use a hold-one-out principle instead. First, we split the estimation window up into a training and validation set. The training set contains the first $W-1$ observations of the estimation window, while the validation set contains its last observation. Next, we perform peLasso with a particular value of λ on the training set to find the optimal combination weights and use these weights on the validation set. This is repeated over a grid of values for the penalty parameter. After this, we use the λ that gave the lowest RMSE in the validation set, in a final peLasso procedure over the full estimation window, which provides combining weights for the out-of-sample period of interest. Note that, contrary to the ex-post analysis, this tuning method provides optimal parameters for each period of interest separately. We use this tuning method in subsection 5.2.

3 Data

For the empirical applications of the peLasso in the next few sections, we use the European Central Bank (ECB) its Survey of Professional Forecasters (SPF). This quarterly held survey started in 1999, coinciding with the launch of the euro, and asks various professional institutions for their expectations regarding macroeconomic variables in the euro area. In particular, the ECB asks for one year ahead forecasts of the HICP (Harmonised Index of Consumer Prices) inflation, the real GDP growth rate and the unemployment level (seasonally adjusted). These expectations are asked for the period one year ahead of the latest available data. Because the survey is held quarterly, this can mean that these "one year ahead" forecasts are sometimes, in fact, only six to twelve months ahead. Additionally, the results of the survey are aggregated and published by taking a simple average, which emphasizes how important this way of combining forecasts is in practice.

Another important feature of the SPF is that the panel of forecasts has frequent gaps, owing to non-response or the entry and exit of institutions. Since all the described methods require a balanced panel, we impute these values with a similar approach as used in Diebold and Shin (2019). First, we only select the most frequent forecasters between 1999Q1 and 2016Q4, which leaves 23, 24 and 25 forecasters for GDP growth, inflation and unemployment, respectively. Then, we apply a linear filter of the form

$$\hat{x}^i_{t+1} - \bar{x}_{t+1} = \theta_i(\hat{x}^i_t - \bar{x}_t) + \epsilon^i_t, \tag{9}$$

where $\bar{x}_t$ is the average of the predictions from the professional institutions that did respond in period t.

This AR(1) process, introduced by Genre et al. (2013), is estimated recursively for each institution. If a forecaster did not respond in the first survey, the recursion cannot start and hence we replace this missing value with the average of the forecasters that did respond. Naturally, replacing the missing starting values in this way biases the estimation of the parameter θ_i downwards. That is why we exclude these observations from the estimation window. To give an example, if an institution has three missing values at surveys 1, 2 and 8, the first two missing values are replaced with the average of the respective surveys and observation 8 is estimated using observations 3 to 7.

Figure 1 shows three histograms of the pairwise correlations between the forecasts of the balanced panel. It clearly shows that the professional forecasts are strongly correlated. This especially

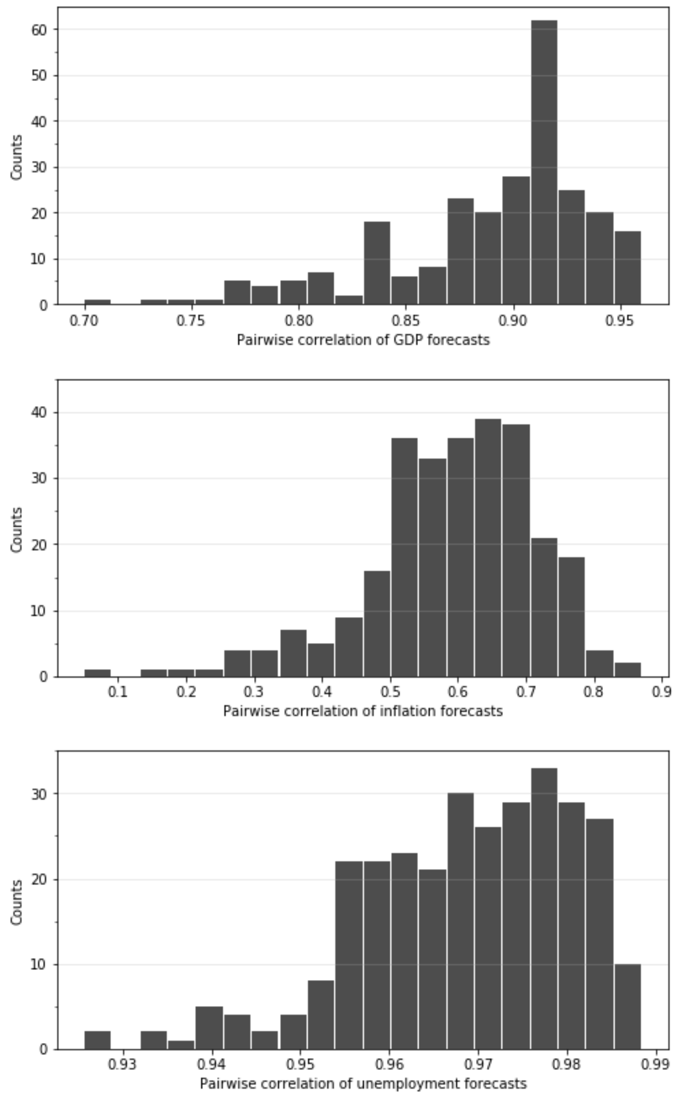

Figure 1: Histograms of pairwise correlations for the GDP growth rate (top), HICP inflation (middle) and unemployment level (bottom) forecasts.

holds for the GDP and unemployment series, which both have a mean correlation of over 0.85. Interestingly, the professionals also tend to agree on which direction the variable is going because there are no negative correlations for any of the series. With these histograms in mind, it is unlikely that the irrepresentable condition holds, which makes Lasso as a selection method questionable.

We obtain the realizations of the macroeconomic variables from Eurostat and plot these in figure 2. Alongside these values, we also graph the predictions of the best and worst forecaster based on the Root Mean Squared Error over the entire sample period 1999Q1-2016Q4. From this figure, it is evident that the series of the GDP growth rate is much more volatile than the other two variables and that this also holds for their respective forecasts. Additionally, even though the best institution performs, on average, better than the worst institution in the full sample, it still has periods where it performs worse than the worst forecaster. This, again, shows that combining forecasts may protect

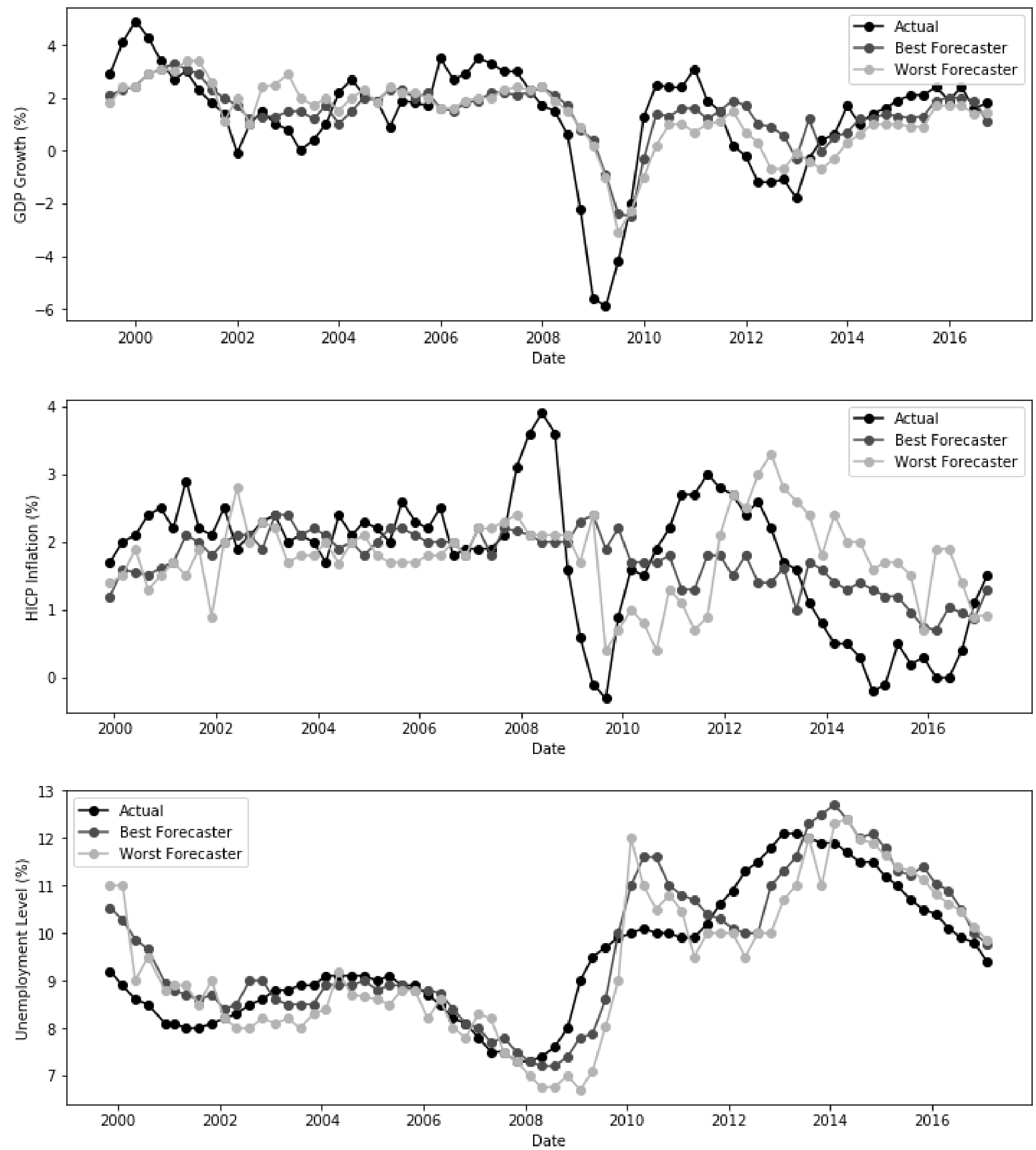

Figure 2: Graphs of the actual values and the forecasts of the best and worst forecaster for the GDP growth rate (top), HICP inflation (middle) and unemployment level (bottom).

a decision-maker from model-specific biases when he or she chooses, a priory, only one model to forecast with. Lastly, it is important to note that the HICP inflation is close to zero for prolonged periods of time, something we come back to in the next section.

4 Ex-post Tuning Results

In this section, we start with the applications of the various peLasso methods to the datasets and tune the parameters ex-post. Each dataset contains forecasts and realizations for 70 periods. Similarly to Diebold and Shin (2019), we use a rolling estimation window of 20 observations to determine the optimal combination weights and use these to make a combined forecast for the next period. However, to make predictions for periods 6-20, we use an expanding estimation window instead. Additionally, due to the inflation being close to zero for prolonged periods, we put a restriction on the selection of forecasters for this dataset, namely that in each period, at least one is chosen. This prevents all combining weights from being set to zero.

First, we perform peLasso in which the trimming method is Lasso and the second step is eRidge, eLasso or an average. To find the optimal tuning parameters, we start with an evenly spaced grid of 200 values in the range [-15, 15] and exponentiate this to (0, 3269017]. Note that when we use eLasso or eRidge as the shrinking method, we perform a search over a two-dimensional grid. The results are given in the upper part of table 1.

The first striking result is that, for each macroeconomic variable, all first step Lasso methods provide the same combined forecasts. This is because shrinkage in the second step for both eLasso and eRidge is strong, such that they become equivalent to taking an average. Secondly, it is optimal to trim most of the forecasters in each period. This is especially true for the inflation forecasts, where on average only 2.72 out of the 24 forecasters remain in each period. Lastly, the RMSE is substantially lower than that of the simple average. For the unemployment level, this difference is also highly significant with a p-value of 0.01. With these results, we confirm the main findings of Diebold and Shin (2019) but extend these to additional datasets besides the GDP growth rate.

Next, we turn to peLasso with Elastic Net as the selection method. Because in some cases we now have to optimize three tuning parameters, we reduce the grid to [0.01, 50]. The results in table 1 show that it still holds that the different second step shrinking methods produce the same forecasts. Harsh trimming remains optimal, but less so compared to the previous first step Lasso results. For example, the average amount of forecasts used went from 4.80 to 6.40 for the GDP growth rate. This is not surprising because, as explained before, Lasso tends to randomly select one predictor out of a correlated group, while Elastic Net can include multiple. More importantly, for all variables, the RMSE also goes down such that for each dataset, it is now lower than that of the best forecaster and significantly lower than the RMSE of the simple average.

After this, we look at the Adaptive Elastic Net. Because this method increases the number of tuning parameters considerably we only perform this with the weighting parameter $\gamma = 1/3$ and only follow this selection method by taking an average over the remaining forecasts[5]. In table 1, we indeed see that the AEN trims a bit harsher than the EN due to punishing smaller coefficients heavier, but this does not lead to a better performance. Of course, this can be due to the fact that we fix the weighting parameter. However, the RMSE does also not improve when we change it to

[5]We set $\gamma = 1/3$ because Diebold and Shin (2019) also try the Adaptive Lasso with this value and find decent results. However, we do not discuss this method.

Table 1: Results of the ex-post analysis with various peLasso methods.

	GDP			Inflation			Unemployment		
Method	RMSE	#	p-val	RMSE	#	p-val	RMSE	#	p-val
peLasso (Lasso, Average)	1.49	4.80	0.13	0.86	2.72	0.09	0.66	3.42	0.01
peLasso (Lasso, eRidge)	1.49	4.80	0.13	0.86	2.72	0.09	0.66	3.42	0.01
peLasso (Lasso, eLasso)	1.49	4.80	0.13	0.86	2.72	0.09	0.66	3.42	0.01
peLasso (EN, Average)	1.44	6.40	0.10	0.82	4.13	0.00	0.62	3.72	0.05
peLasso (EN, eRidge)	1.44	6.40	0.10	0.82	4.13	0.00	0.62	3.72	0.05
peLasso (EN, eLasso)	1.44	6.40	0.10	0.82	4.13	0.00	0.62	3.72	0.05
peLasso (AEN, Average)	1.49	4.30	0.12	0.84	3.25	0.07	0.63	3.30	0.03
peLasso (Puffer, Average)	1.53	10.37	0.21	0.91	12.06	0.49	0.68	6.26	0.15
peLasso (Puffer, eRidge)	1.54	11.21	0.50	0.91	12.24	0.47	0.69	9.34	0.33
peLasso (Puffer, eLasso)	1.53	10.37	0.21	0.91	12.06	0.49	0.68	6.26	0.15
peLasso (GPuffer, Average)	1.52	21.37	0.24	0.89	5.34	0.37	0.66	13.4	0.13
peLasso (GPuffer, eRidge)	1.49	4.80	0.18	0.83	2.29	0.04	0.70	24.6	0.50
peLasso (GPuffer, eLasso)	1.49	4.80	0.13	0.83	2.29	0.04	0.70	24.6	0.50
Comparisons	RMSE	#	p-val	RMSE	#	p-val	RMSE	#	p-val
Best Forecaster	1.45	1	0.19	0.88	1	0.05	0.67	1	0.25
Worst Forecaster	1.81	1	0.99	1.06	1	1.00	0.86	1	1.00
Simple Average	1.54	23	N/A	0.91	23	N/A	0.70	23	N/A

Notes: # is the average number of forecasters used and p-val is the p-value. The p-value is based on the Diebold-Mariano test statistic against the simple average with a one-sided alternative hypothesis and calculated as per Harvey et al. (1997).

2/3, 1 or 4/3 hence we show these results in table A1 in the appendix.

Next, we apply the Puffer transformation before we use Lasso. For every macroeconomic variable, we see, in table 1, a very poor performance. All RMSEs have gone up considerably compared to the other selection methods discussed so far. Upon closer inspection, it is indeed the case that the pairwise correlations between the predictors have been reduced considerably. We illustrate this in figure 3, where we show the histogram of the pairwise correlations between the transformed unemployment forecasts. Similar histograms for the GDP growth rate and inflation can be found in figure A1 in the appendix. Two things stand out from these figures. First off, there are now both positive and negative correlations. Secondly, the absolute values of the pairwise correlations are much lower than those of the untransformed forecasts in figure 1. For example, the mean absolute correlation goes from 0.96 to 0.11 for the unemployment forecasts and from 0.92 to 0.08 for the GDP forecasts. However, there are also many singular values close to zero such that the errors get blown up. This explains why the Puffer transformation is not an improvement despite of the substantial reduction in the pairwise correlations.

To try and solve this issue, we perform the general Puffer transformation with the hard thresholding value $\tau = \sqrt{\frac{P}{T}}$, which is the upper bound suggested by Jia et al. (2015). The general Puffer transformation resolves the issue partly for the GDP growth rate and HICP inflation, such that the

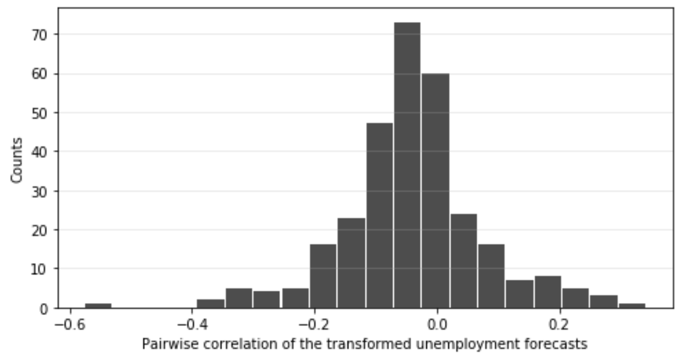

Figure 3: Histogram of pairwise correlations between the transformed unemployment level forecasts.

RMSE is now 1.49 and 0.83, respectively. Nevertheless, the general Puffer transformation still does not improve much on the Elastic Net or Lasso first step selection methods. A potential reason for this is that because $T \approx P$, the hard thresholding value is 1, which is relatively high compared to the singular values. As a consequence, too many values in $\tilde{D}$ are set to zero, a problem that was also found by Brault et al. (2018).

4.1 A closer look at the selected forecasters

Before moving on to the ex-ante tuning methods, we briefly touch on which forecasters are chosen and how many there are selected in each period. As an illustration, figure 4 shows the selected forecasters of the ex-post analysis when we use Lasso as a selection method for the GDP dataset. In this figure, we have given the ex-post best forecaster, over the full sample, the ID 1 and the worst forecaster the ID 23. We show the selected forecasters for the unemployment and inflation forecasts in figure A2 in the appendix.

We can deduce two important things about the nature of the selected set of forecasters from these figures. First off, for each period the selected set tends to contain both ex-post top performers and poorer performers, which Diebold and Shin (2019) call “democratic”. Secondly, it is clear that the forecasters that are selected in one period, are also likely to be selected in the next period. In other words, the selection of forecasters is not independent over time. Interestingly, this also holds for the size of the selected set, which means that even though the selected set of forecasters sometimes changes, the size of the set tends to be the same. These observations further strengthen the motivation behind the hold-one-out tuning method as an alternative to subset-averaging.

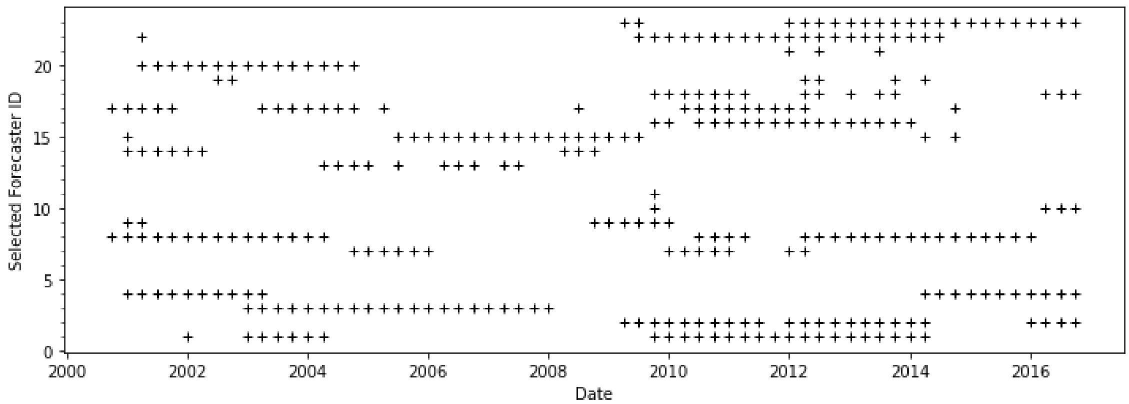

Figure 4: Selected forecasters of the ex-post analysis with first step Lasso selection for the GDP growth rate.

5 Ex-ante Tuning Results

5.1 Subset-averaging Results

The ex-post results have important consequences for subset-averaging. Averaging is still optimal in the second step for all the well-performing peLasso methods in the ex-post analysis, which means that the motivation behind this ex-ante method remains valid. Importantly, this holds for all macroeconomic variables and not only for the GDP growth rate. However, although harsh trimming is still optimal, the Elastic Net trims slightly less hard. Even this slight increase leads to much higher computation times for subset-averaging.

To give an example of this significant increase in computation times, we briefly look back at the number of average forecasters used for the GDP dataset in table 1. Based on the first-step Lasso selection, a reasonable number for $N_{\max}$ is 5. This means that we have to evaluate and sort the performances of ${}_{23}C_1 + {}_{23}C_2 + \cdots + {}_{23}C_5 = 44{,}551$ different subsets for each estimation window. But a proper $N_{\max}$ according to the Elastic Net results is 7, which already gives 390,655 combinations. As a consequence, computation times increase almost nine-fold.

We show the results for $N_{\max} = 5$ and multiple lengths of the estimation window, W, in the upper panel of table 2. We ignore higher subset-sizes as they indeed become infeasible to run even though it would be more appropriate for the GDP dataset[6]. The first observation is that for GDP, the RMSE is very sensitive to the length of the estimation window. For example, while with a window size of 20, subset-averaging performs worse than the simple average, for $W = 2$, the RMSE is much lower than that of the simple average and similar to that of the best forecaster. Nevertheless, for each variable, there is a window length for which subset-averaging produces significantly better forecasts than the simple average and performs as well or better than the best forecaster. Another interesting result is that the average amount of forecasts used is very low, namely around 1.5 for GDP and inflation, and around 1.9 for unemployment. This is in stark contrast with the results in

[6]Typical running times for a fixed window length for $N_{\max} = 5$ are already between one and two hours.

Table 2: Results of the ex-ante subset-averaging method with a maximum subset size of five for different lengths of the estimation window.

	GDP			**Inflation**			**Unemployment**		
Window size	RMSE	#	p-val	RMSE	#	p-val	RMSE	#	p-val
$W = 1$	1.52	1.08	0.36	0.82	1.15	0.00	0.58	1.14	0.00
$W = 2$	1.44	1.52	0.07	0.81	1.51	0.00	0.60	1.62	0.00
$W = 3$	1.46	1.51	0.09	0.82	1.69	0.00	0.59	1.82	0.01
$W = 4$	1.49	1.48	0.22	0.83	1.57	0.00	0.60	1.97	0.01
$W = 5$	1.47	1.49	0.18	0.82	1.51	0.00	0.60	1.95	0.02
$W = 10$	1.49	1.52	0.23	0.83	1.46	0.00	0.64	2.03	0.07
$W = 15$	1.52	1.53	0.34	0.85	1.73	0.01	0.64	2.14	0.11
$W = 20$	1.55	1.51	0.58	0.86	1.78	0.03	0.66	2.35	0.18
$W \leq 20$	1.46	1.50	0.10	0.83	1.55	0.01	0.61	1.88	0.04
Comparisons	RMSE	#	p-val	RMSE	#	p-val	RMSE	#	p-val
Best Forecaster	1.45	1	0.19	0.88	1	0.05	0.67	1	0.25
Worst Forecaster	1.81	1	0.99	1.06	1	1.00	0.86	1	1.00
Simple Average	1.54	23	N/A	0.91	23	N/A	0.70	23	N/A

Notes: # is the average number of forecasters used and p-val is the p-value. The p-value is based on the Diebold-Mariano test statistic against the simple average with a one-sided alternative hypothesis and calculated as per Harvey et al. (1997).

table 1.

Of course, the length of the estimation W has to be chosen ex-ante. To circumvent this problem, Diebold and Shin (2019) suggest a time-varying window. In this case, we pick the subset-average that has performed the best over the estimation windows of length $W \leq W_{\max}$. Consequently, computation times increase even further. We show the results for $W \leq 20$ also in table 2. Notice that for the GDP growth rate, this method performs much better than the simple average and is comparable to the best forecaster. For inflation and unemployment, the results are even better and more significant.

5.2 Hold-one-out Results

Lastly, we turn to the peLasso methods and use the hold-one-out procedure to tune the penalty parameters, which smartly utilizes the dependencies found in the ex-post optimal forecaster selection in figure 4. Additionally, we only consider the peLasso methods with Elastic Net or Lasso as the trimming method and averaging as the second step, because these gave the best results in table 1. Due to this, the grid is only one or two dimensional, which makes the hold-one-out method computationally much cheaper than subset-averaging and also remains feasible for a large number of forecasters.

The results for Lasso and Elastic Net are given in table 3 and table 4, respectively. We obtain

Table 3: Results of the hold-one-out tuning method with Lasso as the trimming method.

	GDP			**Inflation**			**Unemployment**		
Window size	RMSE	#	p-val	RMSE	#	p-val	RMSE	#	p-val
$W = 5$	1.49	5.99	0.09	0.88	4.60	0.04	0.63	5.09	0.04
$W = 6$	1.49	6.06	0.12	0.88	3.94	0.06	0.59	5.50	0.00
$W = 7$	1.53	5.84	0.41	0.86	4.50	0.00	0.62	5.04	0.01
$W = 8$	1.48	5.67	0.15	0.86	4.10	0.01	0.59	4.84	0.00
$W = 9$	1.47	5.59	0.03	0.87	5.31	0.01	0.60	5.30	0.02
$W = 10$	1.47	6.07	0.03	0.83	4.70	0.00	0.61	5.13	0.02
$W = 15$	1.48	5.06	0.08	0.84	3.64	0.00	0.59	5.21	0.00
$W = 20$	1.49	6.09	0.22	0.82	4.56	0.00	0.63	4.76	0.03
$W \leq 20$	1.48	5.83	0.07	0.85	4.33	0.00	0.60	4.93	0.00
Comparisons	RMSE	#	p-val	RMSE	#	p-val	RMSE	#	p-val
Best Forecaster	1.45	1	0.19	0.88	1	0.05	0.67	1	0.25
Worst Forecaster	1.81	1	0.99	1.06	1	1.00	0.86	1	1.00
Simple Average	1.54	23	N/A	0.91	23	N/A	0.70	23	N/A

Notes: # is the average number of forecasters used and p-val is the p-value. The p-value is based on the Diebold-Mariano test statistic against the simple average with a one-sided alternative hypothesis and calculated as per Harvey et al. (1997).

these with a search over the same grid as in the ex- post analysis. Note that we only use estimation windows with at least five observations to have enough training data for the tuning of the penalty parameter.

When we look at table 3, we, again, see a drop in the RMSE compared to the simple average for most variables and estimation windows. This reduction is highly significant for inflation and unemployment, while less so for GDP. If we compare these outcomes with those from subset-averaging, we observe three things. First off, we find that the results for hold-one-out tuning are, in general, less sensitive to the choice of W. Secondly, the performances are quite similar in terms of the RMSE. This holds particularly for the time-varying estimation window, where the differences are around 0.02 for each variable. Lastly, the average amount of forecasters used has increased considerably for each variable.

It is also interesting to briefly look at the behavior of the chosen lambda over time. In figure 5, we plot this time series for the inflation forecasts along with the optimal lambda from the ex-post analysis. The parameter seems to be quite persistent over time, with most values between 0 and 0.5. There are also sometimes sudden jumps to higher values, which indicate that harsher trimming is more optimal. A potential risk of these higher values is that none of the forecasts are selected but this does not happen in our datasets. However, a simple solution for this risk is to pick the best λ that chooses at least one forecaster in the out-of-sample period.

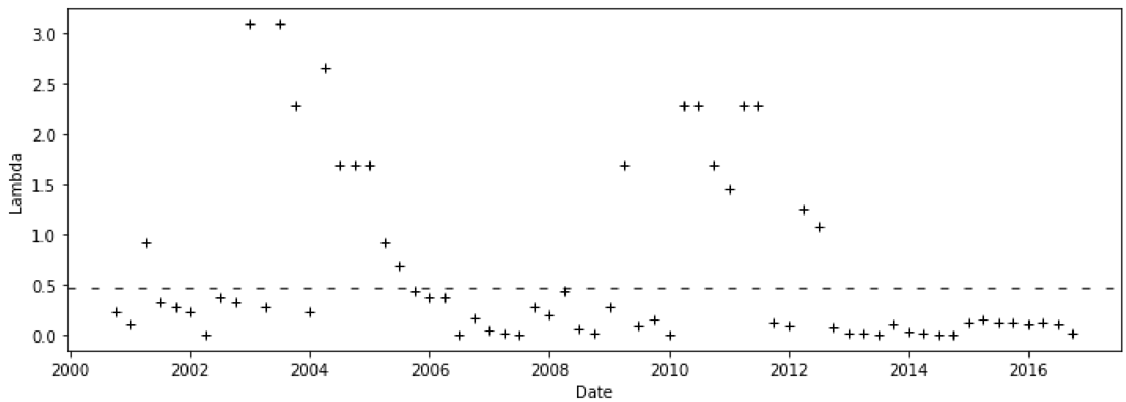

Figure 5: The chosen lambda over time of hold-one-out tuning with Lasso as the selection method for the inflation dataset. The dashed line is the optimal parameter value from the ex-post tuning results.

Table 4: Results of the hold-one-out tuning method with Elastic Net as the trimming method.

	GDP			**Inflation**			**Unemployment**		
Window size	RMSE	#	p-val	RMSE	#	p-val	RMSE	#	p-val
$W = 5$	1.45	7.20	0.01	0.85	4.72	0.02	0.59	5.43	0.02
$W = 6$	1.45	7.28	0.01	0.85	4.52	0.03	0.59	5.52	0.00
$W = 7$	1.47	6.93	0.04	0.83	5.10	0.00	0.60	5.31	0.04
$W = 8$	1.46	6.72	0.00	0.83	5.32	0.00	0.57	5.12	0.00
$W = 9$	1.47	6.82	0.00	0.87	5.41	0.05	0.56	5.08	0.00
$W = 10$	1.45	6.53	0.00	0.81	4.92	0.00	0.56	5.24	0.00
$W = 15$	1.48	7.03	0.11	0.81	4.95	0.00	0.58	5.52	0.00
$W = 20$	1.48	6.99	0.09	0.82	4.88	0.00	0.58	5.03	0.00
$W \leq 20$	1.45	6.89	0.03	0.82	4.82	0.00	0.57	5.11	0.00
Comparisons	RMSE	#	p-val	RMSE	#	p-val	RMSE	#	p-val
Best Forecaster	1.45	1	0.19	0.88	1	0.05	0.67	1	0.25
Worst Forecaster	1.81	1	0.99	1.06	1	1.00	0.86	1	1.00
Simple Average	1.54	23	N/A	0.91	23	N/A	0.70	23	N/A

Notes: # is the average number of forecasters used and p-val is the p-value. The p-value is based on the Diebold-Mariano test statistic against the simple average with a one-sided alternative hypothesis and calculated as per Harvey et al. (1997).

Lastly, for the results of the Elastic Net, in table 4, we see a similar pattern as in table 3. However, in this case, the average amount of forecasters used is even higher, which can, again, be explained by how Elastic Net deals with highly correlated groups of predictors. Overall the performance of the Elastic Net is slightly better than that of Lasso, as was also true in the ex-post tuning results. More importantly, with a flexible window length, peLasso with Elastic Net provides significantly better forecasts than the simple average and comparable or better results than the best forecaster for all three macroeconomic variables.

6 Conclusion

The partially-egalitarian Lasso (peLasso) method introduced and applied by Diebold and Shin (2019) has two shortcomings when it comes to combining professional survey forecasts. First off, professional forecasts are typically highly correlated, a situation Lasso is known to perform poorly in. Secondly, the ex-ante subset-averaging method, which is based on ex-post tuning results that harsh trimming and averaging the remaining forecasts is optimal, is computationally very expensive. Consequently, this research addresses two research questions. Firstly, do selection methods that can deal with high correlations among predictors, improve or change the main results of the ex-post analysis? Secondly, how can we use these new results to construct an ex-ante method that is computationally less expensive?

To answer these questions we first performed peLasso on professional forecasts of the GDP growth rate, inflation and unemployment level, where the tuning parameters were determined ex-post. We found that preconditioning the data, with a Puffer or generalized Puffer transformation, did not provide better results than using the untransformed data directly. However, Elastic Net as the selection method did improve upon Lasso. Importantly, it remained optimal to trim most of the forecast and average the remaining ones. We also found in this ex-post analysis that the selected forecasters were highly dependent over time. We exploited this latter finding to introduce a hold-one-out tuning approach that is completely ex-ante and thus can be used in real-time. More importantly, it produced similar or better results than the subset-averaging approach for the three different datasets of professional forecasts and is computationally much cheaper such that it remains useful in practice. Additionally, when we used this approach with the Elastic Net as the selection method and averaging as the second step in peLasso, it provided significantly better forecast combinations than those of the simple average and similar or much better forecasts than those of the best forecaster.

There are various interesting ways to extend this research. First off, we acknowledge that there are other selection methods that perform well with high correlations that we did not discuss. For example, methods like the Minimax Convex Penalty, Smoothly Clipped Absolute Deviations or preconditioning procedures like the lava transformation. It would thus be interesting to see, whether these change or improve our findings. Secondly, hold-one-out tuning can be susceptible to outliers because it only uses one observation as its validation set. There are more cross-validation approaches that respect time dependencies which are more robust to outliers and perhaps also perform well. For example, nested cross-validation and day forward-chaining. Lastly, so far, the application of peLasso has been limited to forecasts of professional surveys. This is an important limitation because the ex-ante tuning methods are directly derived from these datasets. Even though these techniques can be readily applied to other datasets that do not revolve around professional series, it remains to be seen whether they perform well in those situations.

References

M. Aiolfi and C. A. Favero. Model uncertainty, thick modelling and the predictability of stock returns. *Journal of Forecasting*, 24(4):233–254, 2005.

J. M. Bates and C. W. Granger. The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468, 1969.

V. Brault, C. Lévy-Leduc, A. Mathieu, and A. Jullien. Change-point estimation in the multivariate model taking into account the dependence: Application to the vegetative development of oilseed rape. *Journal of Agricultural, Biological and Environmental Statistics*, 23(3):374–389, 2018.

R. T. Clemen and R. L. Winkler. Combining economic forecasts. *Journal of Business & Economic Statistics*, 4(1):39–46, 1986.

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

G. Elliott. Averaging and the optimal combination of forecasts. *Manuscript, Department of Economics, UCSD*, 2011.

G. Elliott and A. Timmermann. Economic forecasting. *Journal of Economic Literature*, 46(1):3–56, 2008.

V. Genre, G. Kenny, A. Meyler, and A. Timmermann. Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting*, 29(1):108–121, 2013.

C. W. Granger and R. Ramanathan. Improved methods of combining forecasts. *Journal of forecasting*, 3(2):197–204, 1984.

D. Harvey, S. Leybourne, and P. Newbold. Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2):281–291, 1997.

D. F. Hendry and M. P. Clements. Pooling of forecasts. *The Econometrics Journal*, 7(1):1–31, 2004.

A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

J. Jia, K. Rohe, et al. Preconditioning the lasso for sign consistency. *Electronic Journal of Statistics*, 9(1):1150–1172, 2015.

H. Leeb and B. M. Pötscher. Sparse estimators and the oracle property, or the return of hodges' estimator. *Journal of Econometrics*, 142(1):201–211, 2008.

N. Meinshausen, P. Bühlmann, et al. High-dimensional graphs and variable selection with the lasso. *The annals of statistics*, 34(3):1436–1462, 2006.

D. J. Mundfrom, D. G. Shaw, and T. L. Ke. Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, 5(2):159–168, 2005.

P. Newbold and C. W. Granger. Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society: Series A (General)*, 137(2):131–146, 1974.

F. Roccazzella, P. Gambetti, and F. Vrins. Optimal and robust combination of forecasts via constrained optimization and shrinkage. Technical report, LFIN Working Paper Series, 2020/6, 1–2, 2020.

L. J. Tashman. Out-of-sample tests of forecasting accuracy: an analysis and review. *International journal of forecasting*, 16(4):437–450, 2000.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

A. Timmermann. Forecast combinations. *Handbook of economic forecasting*, 1:135–196, 2006.

R. L. Winkler and S. Makridakis. The combination of forecasts. *Journal of the Royal Statistical Society: Series A (General)*, 146(2):150–157, 1983.

C.-H. Zhang et al. Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics*, 38(2):894–942, 2010.

P. Zhao and B. Yu. On model selection consistency of lasso. *Journal of Machine learning research*, 7(Nov):2541–2563, 2006.

H. Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429, 2006.

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

H. Zou and H. H. Zhang. On the adaptive elastic-net with a diverging number of parameters. *Annals of statistics*, 37(4):1733, 2009.

Appendix

Table A1: Results of the ex-post analysis with the Adaptive Elastic Net for various values of the weighting parameter γ.

	GDP			Inflation			Unemployment		
Weighting value	RMSE	#	p-val	RMSE	#	p-val	RMSE	#	p-val
$\gamma = 1/3$	1.49	4.30	0.12	0.84	3.25	0.07	0.63	3.30	0.03
$\gamma = 2/3$	1.49	4.30	0.12	0.84	3.25	0.07	0.63	3.30	0.03
$\gamma = 1$	1.51	4.13	0.25	0.84	3.25	0.07	0.64	3.08	0.04
$\gamma = 4/3$	1.51	4.08	0.18	0.87	2.99	0.08	0.64	3.08	0.04
Comparisons	RMSE	#	p-val	RMSE	#	p-val	RMSE	#	p-val
Best Forecaster	1.45	1	0.19	0.88	1	0.05	0.67	1	0.25
Worst Forecaster	1.81	1	0.99	1.06	1	1.00	0.86	1	1.00
Simple Average	1.54	23	N/A	0.91	23	N/A	0.70	23	N/A

Notes: # is the average number of forecasters used and p-val is the p-value. The p-value is based on the Diebold-Mariano test statistic against the simple average with a one-sided alternative hypothesis and calculated as per Harvey et al. (1997).

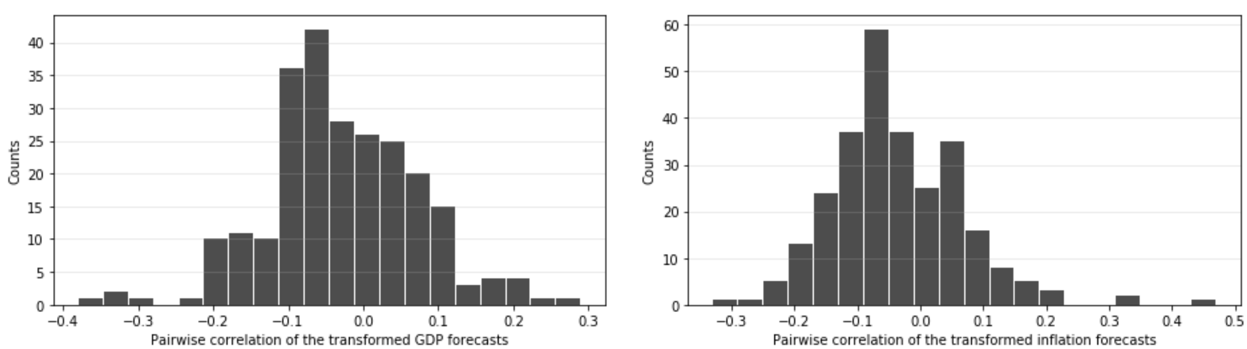

Figure A1: Histograms of pairwise correlations of the transformed GDP growth rate (left) and HICP inflation (right) forecasts.

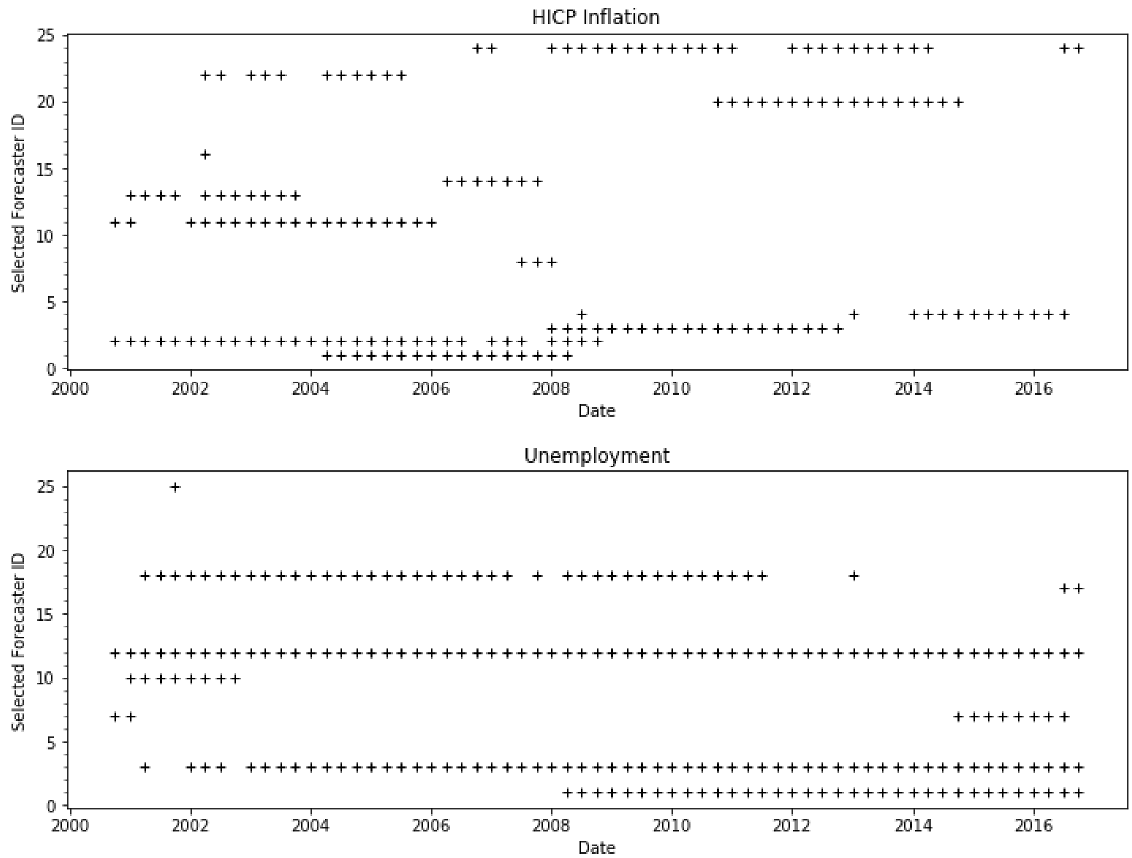

Figure A2: Selected forecasters of the ex-post analysis with first step Lasso selection for HICP inflation (top), and unemployment level (bottom).

Supplementary Code

The .zip file attached to this thesis contains the Python programming code that was used to perform the various methods. The file contains seven .py files which we briefly describe below.

- **Linear_Filter**: This script performs the balancing of the panel with the linear filter of Genre et al. (2013).
- **peLasso(Lasso + optional Puf)**: This script performs peLasso with Lasso as the trimming step and eLasso, eRidge or averaging as the second step. Penalty parameters are optimized ex-post. Optionally, it can first transform the data first with the (generalized) Puffer transformation before applying peLasso.
- **peLasso(EN)**: This script performs peLasso with the Elastic Net as the trimming step and eLasso, eRidge or averaging as the second step. Penalty parameters are optimized ex-post.

- **peLasso(AEN)**: This script performs peLasso with the Adaptive Elastic Net as the trimming step and averaging as the second step. Penalty parameters are optimized ex-post.
- **Subset_Averaging**: This script performs the subset-averaging method for a flexible window length W and $N_{\max}$.
- **Hold_one_out(Lasso)**: This script performs peLasso with Lasso as the trimming step and averaging as the second step. Penalty parameters are optimized with hold-one-out tuning.
- **Hold_one_out(EN)**: This script performs peLasso with the Elastic Net as the trimming step and averaging as the second step. Penalty parameters are optimized with hold-one-out tuning.