ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

(Non-)Linear Regularized Mixtures of Macroeconomic Density Forecasts

Bachelor Thesis Final Report, July 2021

BSc^2 Econometrics/Economics

Thijs Langen

480820

Supervisor: dr. A Tetereva

Second assessor: prof. dr. DJC van Dijk

Abstract

Building on the paper by Diebold, Shin, and Zhang (2021), who propose several regularized mixtures to combine density forecasts, we apply these methods to Eurozone inflation predictive densities retrieved from the Survey of Professional Forecasters (SPF). Beforehand, a simulation discloses the quality of forecasting signals determines the comparative performance of the regularized combinations. For unequal disturbances, Simplex+Ridge with modest penalization strength performs best, which we too find when implementing the mixtures to the SPF. Furthermore, we find regularization effects to shift probability mass to the tails of the predictive densities after the Great Recession and close similarity between Simplex and Best ≤4-Average, in line with Diebold et al. (2021). However, their work only concerns linear aggregation, which especially in this macroeconomic application lacks flexible dispersion of the combined density forecasts (Aastveit, Van Dijk, Mitchel, & Ravazzolo, 2018). Therefore, we allow for this by exploiting the beta-transformed linear pool (BTLP) introduced by Gneiting and Ranjan (2013), extending to Diebold et al. (2021). Such, we improve the performance of the predictive density mixtures for Eurozone inflation by enhancing its calibration. Finally, we discuss Fisher's equation and use it to transform inflation predictive densities to forecast real interest rates.

Table of Contents

1 Introduction

As Granger (1989) asserts, it would have adverse effects on science if econometricians cannot enhance forecasts using new techniques, for instance, by combination methods for which he provided the foundations (Bates & Granger, 1969). Furthermore, an essential shift is occurring from point to density forecasting in the literature, accounting for the probabilistic nature of the target variable (Gneiting & Katzfuss, 2014). In line with the advancement in regularized mixtures for point forecasts presented by Diebold and Shin (2019), a cutting-edge approach would be applying such a combination to density forecasts.

Diebold et al. (2021) effectively lay this out, proposing several regularized mixture methods which are applied to predictive densities for Eurozone inflation, retrieved from the Survey of Professional Forecasters (SPF). Nonetheless, as only linear combinations are used which lack flexible dispersion (Aastveit et al., 2018; Gneiting & Katzfuss, 2014), we build on the work by Diebold et al. (2021) by examining a non-linear mixture, namely the beta-transformed linear pool (BTLP) as proposed by Gneiting and Ranjan (2013). Finally, we provide background on Fisher's equation which we use to transform to real interest rate density forecasts, addressing the "Neo-Fisher" effect as well. Thus, our research question is:

How do regularized mixtures perform under the (beta-transformed) linear pool for Eurozone inflation and real interest rates density forecasts for 1999-2019?

With this research, we expand the current, limited literature on regularized mixtures of density forecasts to non-linear combinations. Although the European Central Bank (ECB) disposes of numerous individual density forecasts to predict inflation, which Diebold et al. (2021) and this paper use, optimal (regularized) mixtures are not conclusively disclosed yet. For such economic applications, the regularized methods of Diebold et al. (2021) are appropriate as often the number of predictors to be combined is relatively low (Diebold & Shin, 2019). However, this too implies a less flexible combined density forecast when using linear aggregation (Aastveit et al., 2018). As the BTLP allows for flexible dispersion and enhanced calibration (Bassetti, Casarin, & Ravazzolo, 2015), implementing this non-linear aggregation technique might provide the ECB with improved inflation forecasts, a major practical contribution of our research.

Starting with applying the regularized mixtures in a simulation, for unequal forecasting signals we find Simplex+Ridge with modest penalization strength to perform best, which we too observe when combining Eurozone inflation density forecasts. Furthermore, we encounter regularization effects after the Great Recession, and Simplex showing strong resemblance with Best $\leq$4-Average, in line with Diebold et al. (2021). Moving towards the BTLP, for the simulation, we observe the BTLP to cause increased log scores for multiple regularized mixtures. Still, we might prefer using "One-Time" Estimation to avoid moving towards the wrong model during the Alternating Estimation, where we iteratively refit weights and BTLP parameters. This phenomenon appears at our empirical application to Eurozone inflation as well, although we only see a small and insignificant decrease of log score for "One-Time" Estimation. Here, PIT series show calibration vastly improves, confirming the advantages of the flexible dispersion brought about by the BTLP compared to linear aggregation. Finally, for each aggregation technique, we transform to real interest rate density forecasts using Fisher's equation, which comes down to a location shift and mirror of the "original" predictive density. Such, we can directly construct density forecasts for another macroeconomic variable. Furthermore, we demonstrate even simpler conversion when using the "Neo-Fisher" equation, introduced by Uribe (2018).

This paper is structured as follows. First, we discuss the literature on combining (density) forecasts (section 2), emphasizing non-linear aggregation techniques. Second, we introduce the SPF and Eurozone macroeconomic data (section 2), after which the regularized mixture methods are provided in the methodology (section 3). Moreover, in this part, we describe the simulation procedure and application of the methods to Eurozone inflation. Furthermore, we demonstrate the Fisher transformation and implementation of the BTLP. Consequently, we present our results (section 5) for both aggregation techniques. Finally, we state a conclusion and identify some points of discussion (section 6).

2 Literature

Although density forecasts disclose more information than interval or point forecasts, the latter is the most used in practice and academics. This could be explained by the difficulties in assumptions, interpretation and computation emerging when assessing predictive densities (Diebold, 2000). Hence, we start with an important class of point forecasting covering the combination of multiple predictions.

2.1 Literature Combined Point Forecasting

Bates and Granger (1969) offered the basis for this by considering linear combinations, using equal or optimal weights in the sense of minimizing mean-square errors. From then, scholars embraced this approach as a fruitful way to increase predictive performance and many extensions were investigated. For instance, by iteratively evaluating the performance of forecasters and changing allocation of importance, weights can be made time-varying (Aastveit et al., 2018; Timmermann, 2006). Moreover, practical applications abound - Clark and McCracken (2010), for instance, argue combining forecasts to reduce instability in macroeconomic VAR models. Additionally, Genre, Kenny, Meyler, and Timmermann (2013) employ various forecast combinations to predict Eurozone macroeconomic variables, which Diebold and Shin (2019) extend by addressing penalized estimation problems, both using the same SPF dataset as this paper. This relates to the work by Bayer (2017), in which Value-at-Risk forecasts for the Dow Jones index are combined similarly. Noteworthy is the "equal weights puzzle", which is the stylized fact of superior equal weights performance due to the estimation error arising when optimizing a weighting schedule (Aastveit et al., 2018; Diebold & Shin, 2019; Genre et al., 2013).

2.2 Literature Density Forecasting

Nonetheless, the above-mentioned account only considers single-valued predictions, while Gneiting and Katzfuss (2014) acknowledge and endorse a shift towards predictive densities to specify the uncertainty featured by any forecast. This offers broad applications in meteorology, health care, economics and finance (Gneiting & Katzfuss, 2014). For instance, Diebold, Gunther, and Tay (1998) construct density forecasts for the S&P 500 using a Normal and Student's t-distribution and a Moving Average and GARCH approach to iteratively estimate mean and variance. Consequently, evaluation is done through using visual assessment of the histograms and correlograms of the PIT series, which should be *iid* uniformly distributed under ideal forecasting, a widespread procedure within the literature (Diebold et al., 1998; Gneiting & Katzfuss, 2014; Raftery, Gneiting, Balabdaoui, & Polakowski, 2005). Moreover, scoring rules, which map the predictive density and realization to one numerical value, could be used to assess or compare distributional forecast performance. Prominent examples are the Logarithmic Score (LS), Brier Score (BS), Quadratic Score (QS) and Continuously Ranked Probability Score (CRPS) (Bröcker & Smith, 2007; Gneiting & Katzfuss, 2014; Gneiting & Raftery, 2007).

2.3 Literature Combined Density Forecasting

Similarly to point forecasting, there exists a category within density forecasting considering combinations of predictive densities, initially implementing linear mixtures denoted as the "linear opinion pool" (Geweke & Amisano, 2011; Hall & Mitchell, 2007; Wallis, 2005). Nonetheless, Gneiting and Katzfuss (2014) state this class lacks flexible dispersion such that general methods should be used to improve prediction performance (Aastveit et al., 2018; Gneiting & Katzfuss, 2014). Therefore, various non-linear aggregation methods have been proposed which are described below in Table 1, including their advantages and disadvantages.

Table 1

Overview of non-linear aggregation methods for combining density forecasts

Non-linear aggregation method	**Description**	**Advantages**	**Disadvantages**	**References**
Generalized Linear Pool (GLP)	Adjusts individual forecasters with some function $g(\cdot)$, after which the linear combination is implemented and evaluated at $g^{-1}(\cdot)$.	Improved calibration (compared to LP).	Poor flexible dispersion in some cases, need to specify function $g(\cdot)$ (misspecification bias).	Aastveit et al. (2018); Gneiting and Ranjan (2013)
Spread-adjusted Linear Pool (SLP)	Scales the observations and weights by the parameter c, after which the linear combination is implemented.	Improved calibration (compared to LP).	Poor flexible dispersion, need to specify/estimate c (estimation error).	Aastveit et al. (2018); Glahn et al. (2009); Gneiting and Ranjan (2013)
Beta-Transformed Linear Pool (BTLP)	Implements the linear combination, which is adjusted by a linear combination of the individual predictive CDF's evaluated at the pdf of the beta distribution ($b_{\alpha,\beta}$).	Improved calibration (compared to LP), good flexible dispersion.	Need to specify/estimate α and β (estimation error).	Aastveit et al. (2018); Bassetti et al. (2015); Gneiting and Ranjan (2013)

Following Table 1, the most advantageous method seems to be the beta-transformed linear pool (BTLP) proposed by Gneiting and Ranjan (2013) and used as well by Bassetti et al. (2015). Therefore, we will implement this non-linear aggregation technique in our research, as discussed later in the methodology (in particular, subsection 4.5).

Furthermore, the weighting scheme could be specified as equal weights or through optimizing a certain objective (Aastveit et al., 2018). In the latter case, a common approach is to minimize the Kullback-Leibler Information Criterion (KLIC) divergence between forecasted and true density forecast, which by construction comes down to minimizing the average logarithmic score (Aastveit et al., 2018; Geweke & Amisano, 2011). Hall and Mitchell (2007) discuss the same method and add the possibility of minimizing a goodness-of-fit test statistic, such as the Kolmogorov-Smirnov or Berkowitz LR-test (Hall & Mitchell, 2007). Lately, Diebold et al. (2021) implemented regularized mixtures to combine density forecasts, on which this paper builds. They examine various methods (Simplex, Simplex+Ridge, Simplex+Entropy and Subset Averaging) and apply them to Eurozone macroeconomic variables, building on Diebold and Shin (2019) and Genre et al. (2013). Similar to this paper is the work of Cumings-Menon and Shin (2020), who combine densities using regularized Wasserstein distance.

An overview of the discussed literature on the various types of forecasting is presented in Appendix A.

2.4 Economic Background: Fisher's Equation

While Diebold et al. (2021) link inflation and real interest rate with a simple transformation following the Fisher equation, we discuss this relationship and its validity more broadly by assessing empirical evidence. Fama (1975) identifies the Fisher effect by using bond market efficiency and the nominal interest rate this market sets. Consistent with this finding is the work by Booth and Ciner (2001), who find a one-to-one relation between Eurocurrency rates and inflation in the long run. Consequently, Mishkin (1991) emphasizes the Fisher effect to only hold in the long run, with no link between inflation and nominal interest rate in the short run. However, Koustas and Serletis (1999) show the equation to not even hold for the long term for many European countries. Even more, Jaffe and Mandelker (1976) attempt to apply the relation to stock market returns and inflation, but cannot unveil the Fisher effect for their data. In fact, they find a negative effect between inflation and market returns for some periods (Jaffe & Mandelker, 1976). Accordingly, a recent finding is the "Neo-Fisher effect", demonstrated by Uribe (2018), distinguishing between temporary and permanent changes in nominal interest rate. While in the temporary case the equation holds, for permanent increments of nominal interest rate inflation increases too, so opposite to what Fisher argued (Uribe, 2018).

3 Data

In 1999, the ECB started with the Survey of Professional Forecasters (SPF) to construct forecasts of inflation, GDP growth and unemployment in the Eurozone. As its name suggests, experts with major forecasting experience provide their outlook for these macroeconomic variables (Genre et al., 2013). This is done quarterly, both by stating point forecasts as well as assigning probabilities to outcome intervals, the latter used in this research. Furthermore, several forecast horizons are specified, such as a prediction one year ahead of the latest moment realized inflation is published. Following Diebold et al. (2021), we use the data with this horizon specification on predictive densities for inflation ranging from 1999Q1 to 2019Q3, so including 83 quarters. The dataset is publicly available and retrieved from the ECB website.[1]

Moreover, the target variable of the inflation prediction is defined as the percentage change in the Harmonised Index of Consumer Prices (HICP) for the Euro area. Naturally, we include 83 realizations, ranging from December 1999 to June 2020, which are publicly available too and retrieved from the website of the Federal Reserve Bank of St. Louis (FRED).[2] As the inflation

[1] `https://www.ecb.europa.eu/stats/ecb_surveys/survey_of_professional_forecasters/html/index.en.html`

[2] `https://fred.stlouisfed.org/series/CP0000EZ19M086NEST` (Although this is retrieved from an American source, it does concern Eurozone data)

forecast is made for the period starting in the quarter of the survey and ending a year later in the month the latest HICP data was available, we define the HICP realized percentage change likewise (Diebold et al., 2021; Genre et al., 2013).

Finally, interest rates are used to transform inflation density forecasts to construct predictive densities for real interest rate by Fisher's equation. Therefore, we retrieve 3-Month Eurozone Interbank Rates, which are too publicly available and found at the FRED website. [3]

4 Methodology

In the following sections, we describe our methodology by starting to discuss the fundamentals of combining density forecasts (subsection 4.1) and regularized mixtures (subsection 4.2). Subsequently, we describe our simulation (subsection 4.3), empirical application to the Eurozone inflation forecasts (subsection 4.4) and extension to the BTLP (subsection 4.5). All implementation will be done in R, where we use the package `BB` (Varadhan & Gilbert, 2009) to optimize weights, and for which the programming code is attached to this paper.

4.1 Fundamentals Combining Density Forecasts

Starting with notational matters, we denote our target variable (HICP percentage change) as y_t for $t = 1, ..., T$. These observations are realised from a certain DGP, which can be characterized with f_t and can be predicted with p_t for $t = 1, ..., T$. As it is possible to have multiple predictors, we define $p_{k,t}$ as the k^{th} density forecast for $k = 1, ..., K$ and $t = 1, ..., T$ with CDF $P_{k,t}$. Such, we can create a combined predictive density $\tilde{p}_t$ for $t = 1, ..., T$. Hence, in the case of a linear mixture we have $\tilde{p}_t = \sum_{k=1}^{K} \mathrm{w}_k p_{k,t}$, so including a weighting vector $\mathbf{w} = (\mathrm{w}_1, ..., \mathrm{w}_K)'$. Therefore, for a Simple Average we have $\mathrm{w}_k = 1/K$ for $k = 1, ..., K$ (Aastveit et al., 2018). Still, we could also estimate these weights by optimizing a certain objective, for which we use the log score defined as:

$$LS(p_{k,t}; y_t) = -\log\big(p_{k,t}(y_t)\big), \tag{1}$$

reducing to $-\log\big(\tilde{p}_t(y_t)\big)$ for the combined density forecast. As we use linear aggregation and minimize the sum of this log score over the period, we have the estimation problem:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_{t=1}^{T} -\log\Big(\sum_{k=1}^{K} \mathrm{w}_k p_{k,t}(y_t)\Big), \tag{2}$$

in accordance with Hall and Mitchell (2007) and Geweke and Amisano (2011).

[3] https://fred.stlouisfed.org/series/IR3TIB01EZQ156N (Although this is retrieved from an American source, it does concern Eurozone data)

4.2 Regularized Mixtures for Combining Density Forecasts

We expand the basic estimation problem in Equation 2 with various constraints and penalties (including penalty parameter λ). This results in the subsequent regularized mixture methods following Diebold et al. (2021):

1. **Simplex**

$$\begin{aligned} \mathbf{w}^* = \arg\min_{\mathbf{w}} \quad & \sum_{t=1}^{T} -\log\Big(\sum_{k=1}^{K} \mathrm{w}_k p_{k,t}(y_t)\Big) \\ \text{s.t.} \quad & \mathrm{w}_k \in (0,1) \text{ for } k = 1, ..., K \text{ and } \sum_{k=1}^{K} \mathrm{w}_k = 1, \end{aligned} \tag{3}$$

   such that we add non-negativity and sum-to-one constraints, which are essential for combining density forecasts to arrive at a proper distribution (Aastveit et al., 2018; Yao, Vehtari, Simpson, & Gelman, 2018), contrary to point forecasts for which it is only recommended to evade multicollinearity. Diebold et al. (2021) remark this method already provides some regularization: since the Simplex constraints demarcate a restricted parameter space, it is a subclass of LASSO.

2. **Simplex + Ridge**

$$\begin{aligned} \mathbf{w}^* = \arg\min_{\mathbf{w}} \quad & \left(\sum_{t=1}^{T} -\log\Big(\sum_{k=1}^{K} \mathrm{w}_k p_{k,t}(y_t)\Big) \;+\; \lambda\Big(\sum_{k=1}^{K}\Big(\mathrm{w}_k - \frac{1}{K}\Big)^2\Big) \right) \\ \text{s.t.} \quad & \mathrm{w}_k \in [0,1] \text{ for } k = 1, ..., K \text{ and } \sum_{k=1}^{K} \mathrm{w}_k = 1, \end{aligned} \tag{4}$$

   where λ denotes the penalization parameter. Such, additionally to Equation 3, weights are shrunk towards $1/K$ to avoid potential parsimony caused by Simplex alone (Diebold et al., 2021).

3. **Simplex + Entropy**

$$\begin{aligned} \mathbf{w}^* = \arg\min_{\mathbf{w}} \quad & \left(\sum_{t=1}^{T} -\log\Big(\sum_{k=1}^{K} \mathrm{w}_k p_{k,t}(y_t)\Big) \;+\; \lambda\Big(-\sum_{k=1}^{K} \log(\mathrm{w}_k)\Big) \right) \\ \text{s.t.} \quad & \mathrm{w}_k \in (0,1) \text{ for } k = 1, ..., K \text{ and } \sum_{k=1}^{K} \mathrm{w}_k = 1, \end{aligned} \tag{5}$$

   where λ denotes the penalization parameter. Similar to Simplex+Ridge, forecasters which would receive no or very low importance for Simplex, now are assigned a weight forced in the direction of $1/K$. This is because when $\mathrm{w}_k \to 0$, $\log(\mathrm{w}_k) \to -\infty$ such that often, even for modest penalization strength, all forecasters are included (Diebold et al., 2021).

4. **Subset averaging**

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \quad \Bigg(\sum_{t=1}^{T} -\log\Big(\sum_{k=1}^{K} \mathrm{w}_k p_{k,t}(y_t)\Big) \; + \; \lambda\Big(\sum_{k=1}^{K}\Big(\mathrm{w}_k - \frac{1}{\delta(\mathbf{w})}\Big)^2\Big) \Bigg)$$
$$\text{with } \lambda \to \infty \tag{6}$$
$$\text{s.t.} \quad \mathrm{w}_k \in [0,1] \text{ for } k = 1, ..., K \text{ and } \sum_{k=1}^{K} \mathrm{w}_k = 1,$$

where λ denotes the penalization parameter and $\delta(\mathbf{w})$ the number of elements in $\mathbf{w}$ which are non-zero. As $\lambda \to \infty$, this comes down to using the best performing average at $t = 1, ..., T$ (if λ is unrestricted, we have the simplex-constrained partially egalitarian ridge problem). Two classes of subset averaging could be distinguished. First, "Best N-Average", for which each time the best N forecasters are used and averaged. Second, "Best $\leq N_{max}$-Average", for which each time the best $\leq N_{max}$ forecasters are used and averaged. Such, if we have a set of K forecasts, we consider ${}_K C_N$ possible averages in the first case and $\sum_{i=1}^{N_{max}} {}_K C_i$ in the second (Diebold & Shin, 2019; Diebold et al., 2021).

4.3 Methodology Monte Carlo Simulation

Consequently, an initial assessment of these regularized mixtures is done through Monte Carlo simulation. For Subset Averaging, only Best 4-Average and Best ≤4-Average are used, in line with the empirical application of Diebold et al. (2021). For the simulation, the next (known) DGP and procedure following Diebold et al. (2021) is used:

$$\begin{aligned} y_t &= x_t + \sigma_y e_t, \qquad & e_t &\sim \text{iid}\, \mathcal{N}(0,1) \\ x_t &= \phi_x x_{t-1} + \sigma_x v_t, \qquad & v_t &\sim \text{iid}\, \mathcal{N}(0,1), \end{aligned} \tag{7}$$

for $t = 1, ..., T$ and y being the target variable of our forecast. Furthermore, e_t and v_s are independent for all $t, s = 1, ..., T$. We have K individual forecasters, which are specified as:

$$z_{k,t} = x_t + \sigma_{z,k}\eta_{k,t}, \quad \eta_{k,t} \sim \text{iid}\, \mathcal{N}(0,1), \tag{8}$$

for $k = 1, ..., K$, $t = 1, ..., T$ and where $\eta_{k,t}$ and $\eta_{l,s}$ are independent for all $t, s = 1, ..., T$ and $k, l = 1, ..., K$ (unless, of course, when $t = s$ and $k = l$). Moreover, we use 1-step ahead predictive densities following a Normal distribution for which the mean is unknown, which we assess using the individual forecasters. Nonetheless, the standard deviation σ_y is assumed to be known. Therefore, the individual density forecast results in:

$$p_{k,t}(y_{t+1}) = \mathcal{N}(\phi z_{k,t},\ \sigma_y^2). \tag{9}$$

The DGP is split into two types (DGP 1 and DGP 2) by using two specifications for $\sigma_{z,k}$,

1. $\sigma_{z,k} = 1$ for $k = 1, ..., K$ (DGP 1).
2. $\sigma_{z,k} = 1$ for $k = 1, ..., \frac{K}{2}$ and $\sigma_{z,k} = 5$ for $k = \frac{K}{2} + 1, ..., K$ (DGP 2).

This is done to distinct between forecast signals, as in the first case disturbances are similar, while in the second not such that the weighting scheme should correct for this. Therefore, we might expect the relative performance of Simple Average compared to Simplex will be high for equal disturbances (DGP 1) and lower when disturbances become more unequal (DGP 2). However, both types involve the parameter specifications $\phi_x = 0.9$, $\sigma_x = 1$ and $\sigma_y = 0.5$ combined with $K = T = 20$. After the data is generated, optimal weights are estimated such that we can construct a 1-step-ahead combined density forecast, so for the observation at $t = T + 1 = 21$. Furthermore, various penalization strengths (λ) are implemented for Simplex+Ridge and Simplex+Entropy, 20 for each one. For Simplex+Ridge, we use 10 equispaced points in [1e-15, 10] and in [15, 10 000]. For Simplex+Entropy, we use 10 equispaced points in [1e-15, 0.2] and in [0.3, 20]. Finally, for all methods, we report the average log scores and the average number of used predictors (number of non-zero weights) of the combined density forecasts for the 10 000 Monte Carlo replications. Moreover, a graph of the average log scores with respect to the investigated penalization strengths is constructed such that the ex-post optimal penalization strength (λ^*) for Simplex+Ridge and Simplex+Entropy can be disclosed (Diebold et al., 2021).

4.4 Methodology Empirical Application

After that, the regularized mixture methods examined in the simulation are applied to the SPF density forecasts for inflation (see section 3). Although the dataset does not come in a continuous pdf as probabilities are assigned to bins, the above-mentioned accounts have a straightforward discrete analogue. For instance, as $p_{k,t}(y_t)$ comes down to the probability assigned to the bin in which y_t falls, the log score is easy to determine. Nonetheless, before the regularized mixtures are applied, three data issues need to be resolved in accordance with Diebold et al. (2021):

1. **Bin classification inconsistency.** Over the sample period, bin definitions did not remain constant. Whereas bin width has always been 0.5 (HICP inflation in %), there occurred some changes at the tail bins of the predictive density, as can be seen below in Table 2.

Table 2

Bin definitions predictive densities Eurozone inflation ECB-SPF, 1999Q1 - 2019Q3

Interval (in %) \Period	1999Q1 - 2000Q3	2000Q4	2001Q1 - 2008Q2	2008Q3 - 2009Q1	2009Q2 - 2009Q4	2010Q1 - 2019Q3	**Used bins**
-2.5 to -2.0					x		
-2.0 to -1.5					x		
-1.5 to -1.0					x	x	
-1.0 to -0.5					x	x	x
-0.5 to 0	x	x	x	x	x	x	x
0 to 0.5	x	x	x	x	x	x	x
0.5 to 1.0	x	x	x	x	x	x	x
1.0 to 1.5	x	x	x	x	x	x	x
1.5 to 2.0	x	x	x	x	x	x	x
2.0 to 2.5	x	x	x	x	x	x	x
2.5 to 3.0	x	x	x	x	x	x	x
3.0 to 3.5	x	x	x	x	x	x	x
3.5 to 4.0	x	x	x	x	x	x	x
>4.0		x		x	x	x	x

To achieve a consistent bin definition for our entire dataset, we use 11 bins of width 0.5, starting from the interval $(-\infty, -0.5]$ and ending at $(4, +\infty]$ as displayed in Table 2. Thus, we add some extreme bins (with zero loadings) to the bounds of the original density forecasts for 1999Q1-2009Q1 and merge bins at the left-hand side of the predictive density for 2009Q2-2019Q3 (Diebold et al., 2021).

2. **Forecaster group discontinuity.** Although there are over 103 distinct forecasters in the dataset, none of them replies to the survey each period, resulting in an unbalanced panel. Therefore, we first exclude forecasters who do not appear in the survey pool for four successive quarters, in line with Genre et al. (2013) and Diebold et al. (2021). Still, many gaps remain unfilled. We resolve this by first replacing the empty forecasts at $t = 1$ (1999Q1) with the average of the non-missing predictive densities. Subsequently, we compute the Ranked Score (RS) for each predictive density $p_{k,t}$, in the discrete case defined as:
$$RS(p_{k,t}; y_t) = \sum_{h=1}^{H} (P_{k,t,h} - O_h)^2, \tag{10}$$
where $P_{k,t,h}$ is the hth bin of the CDF of $p_{k,t}$ for $h = 1, ..., H$, so $P_{k,t,h} = \sum_{j=1}^{h} p_{k,t,j}$. Furthermore, $O_h = \sum_{j=1}^{h} \mathbb{1}_j(y_t)$, where $\mathbb{1}_j(y_t)$ equals 1 if y_t falls in the jth bin and equals 0 else (Weigel, Liniger, & Appenzeller, 2007). Based on the RS, we order our forecasters from high to low predictive performance and arrange them in five groups. For the second survey at $t = 2$ (1999Q2), the empty predictive density will be set to the average of forecasters who are present in its group. After that, the RS will be recalculated and forecasters will be regrouped such that we can fill the gaps of the next survey. This procedure will be repeated till the final survey at $t = T = 83$ (2019Q3).

3. **Log score zero problem.** Finally, a problem occurs when an inflation observation falls

into a bin that was assigned zero probability as in this case the log score returns $-\infty$ (see Equation 1). To resolve, we assign a probability of 0.01 (1%) to such a bin. Naturally, as the integral over the predictive density (so the sum of the bins) should still be 1 (100%), we divide 0.01 by the number of non-zero bins and subtract this amount from these bins.

After completing these refinements of the dataset, we implement the regularized mixtures of subsection 4.2. For Simplex+Ridge and Simplex+Entropy we try the same penalization strengths (λ) as used in the simulation (see subsection 4.3), to disclose the optimal penalty parameter (λ^*). Moreover, for all methods we include a uniform density forecast, so a predictor with loading 1/11 in each of the 11 bins. While Diebold et al. (2021) do not include this "constant" forecaster for the Simple Average, we do as in this case a more forthright comparison can be made with the regularization methods.

In line with our simulation and Diebold et al. (2021), we use a moving window of 20 quarters to estimate the optimal weights, which we use to construct 1-step-ahead mixed density forecasts. The observations ranging from 1999Q1 to 2000Q4 (8 observations, $t = 1, ..., 8$) are set as the burn-in sample, so the evaluation period ranges from 2001Q1 to 2019Q3 (75 observations, $t = 9, ..., 83$). In line with Diebold and Shin (2019), who perform similar research on 1-step ahead combined forecasts for the same dataset, we compute the optimal weights to construct combined predictive densities for the observations at $t = 9, ..., 20$ by using all realizations made from $t = 1$ until that observation. After $t = 20$, we use the entire 20-quarter moving window.

For evaluation, we again report the average log scores (including the graph with respect to penalization strength) and the average number of used predictors for all methods. Furthermore, to assess calibration, we check the histograms of the PIT series, defined as $PIT_t = \int_{-\infty}^{y_t} \tilde{p}_t(s)\, ds = \tilde{P}_t(y_t)$, which should be *iid* uniformly distributed under ideal forecasting (Gneiting & Katzfuss, 2014), also in the discrete case (Czado, Gneiting, & Held, 2009). In addition, we create heat maps of differences between densities for regularized mixtures and Simple Average, and between Simplex and Best $\leq$4-Average.

Finally, we transform the inflation distributions to real interest rate density forecasts by using Fisher's equation ($r_t = i_t - \pi_t$) such that $p_t^r(\cdot) = i_t - p_t^\pi(\cdot)$, where p_t^r and p_t^π respectively denote real interest rate and inflation density forecasts for $t = 1, ..., T$. Hence, we implement a shift in location and mirror the "original" inflation density forecasts (Diebold et al., 2021). Nonetheless, it would be irrelevant to display log scores, as it is invariant for such transformation, nor PIT histograms as these are mere mirrors of the ones for inflation. Both statements are proven in Appendix B. As discussed in our economic background (subsection 2.4), the Fisher equation does not rigorously hold in reality, decreasing the practical value of this transformation. Nonetheless, for theoretical purposes it is instructive to perform such conversion of density forecasts, also to see the consequences of using the "Neo-Fisher effect" presented by Uribe (2018). In this case, the transformation will be $p_t^r(\cdot) = i_t + p_t^\pi(\cdot)$, so only a location shift remains. Amalgamated with the fact that log scores remain invariant and PIT histograms are even the same (proof in Appendix B), the conversion from inflation to real interest rate forecasting will only be simpler should the "Neo-Fisher effect" be properly identified. Therefore, to avoid tediousness, we refrain from including graphs for this effect.

4.5 Extension to Beta-Transformed Linear Pool

Extending the research by Diebold et al. (2021), we implement the beta-transformed linear pool (BTLP) as proposed by Gneiting and Ranjan (2013) to allow for flexible dispersion, which is a key advantage compared to linear combinations (Aastveit et al., 2018; Gneiting & Katzfuss, 2014) and to the other non-linear aggregation techniques discussed in Table 1 (in subsection 2.3).

To combine predictive cumulative density functions (CDF's) $P_{k,t}$, Gneiting and Ranjan (2013) define the BTLP as:

$$\tilde{P}_t = B_{\alpha,\beta}\left(\sum_{k=1}^{K} \mathrm{w}_{k,t} P_{k,t}\right), \tag{11}$$

where $B_{\alpha,\beta}(\cdot)$ denotes the CDF of the beta distribution, elaborated on in Appendix C. As $B_{\alpha,\beta}(\cdot)$ is a monotone mapping $[0,1] \to [0,1]$ depending on α and β, the intuition behind this class is we correct the CDF combination without making it invalid. Furthermore, equal weights do not mean a Simple Average of the forecasters, compared to linear combinations. Nonetheless, when $\alpha = \beta = 1$, the beta distribution comes down to an uniform distribution, so $B_{\alpha,\beta}(x) = x$. Hence, in this case, the combination method for our forecasters comes down to linear aggregation, showing that the linear pool (LP) is a subclass of the BTLP (Gneiting & Ranjan, 2013). Furthermore, to come to a combined probability density function in accordance with our existing methodology, we take the derivative of the combined CDF of Equation 11 which leads to:

$$\tilde{p}_t = \underbrace{\left(\sum_{k=1}^{K} \mathrm{w}_{k,t} p_{k,t}\right)}_{\text{Linear combination}} \cdot \underbrace{b_{\alpha,\beta}\left(\sum_{k=1}^{K} \mathrm{w}_{k,t} P_{k,t}\right)}_{\text{BTLP ``Correction''}}, \tag{12}$$

where $b_{\alpha,\beta}(\cdot)$ denotes the PDF of the beta distribution and for which the derivation can be found in Appendix C, in line with Gneiting and Ranjan (2013). To estimate parameters α and β, we pursue an analogue approach to optimizing the weights, so by minimizing the log score. We commence by assessing this extended optimal weighting problem in a quite straightforward way as follows. First, for all discussed (regularization) methods, we start with the optimal weights for linear aggregation and estimate the optimal α by minimizing the log score, keeping weights at the initial optimal LP weights and $\beta = 1$:

$$\alpha^* = \arg\min_{\alpha} \sum_{t=1}^{T} -\log\Bigg(\Big(\sum_{k=1}^{K} \mathrm{w}^*_{k,t} p_{k,t}(y_t)\Big) b_{\alpha,1}\Big(\sum_{k=1}^{K} \mathrm{w}^*_{k,t} P_{k,t}(y_t)\Big)\Bigg). \tag{13}$$

Second, the optimal β will be estimated, keeping weights at the initial optimal LP weights and $\alpha = \alpha^*$:

$$\beta^* = \arg\min_{\beta} \sum_{t=1}^{T} -\log\Bigg(\Big(\sum_{k=1}^{K} \mathrm{w}^*_{k,t} p_{k,t}(y_t)\Big) b_{\alpha^*,\beta}\Big(\sum_{k=1}^{K} \mathrm{w}^*_{k,t} P_{k,t}(y_t)\Big)\Bigg). \tag{14}$$

Alternative to this "One-Time" Estimation, we could refine the estimation of this extended problem by using the optimal α and β to re-estimate the optimal weights for the beta-transformed linear pool in Equation 12. Consequently, we too re-compute the optimal α and β by minimizing the log score as in Equations 13 and 14. We repeat this iterative estimation procedure 5 times (for simplicity and computational reasons), such that we implement an Alternating Estimation approach reflected by the algorithm presented in Appendix C. Hence, we add 5 extra iterations to the simple "One-Time" Estimation. Implementation of all regularization methods proceeds similarly as explained earlier, so by adding the relevant penalty and constraints in the estimation problem (subsection 4.2). However, for Simplex+Ridge and Simplex+Entropy, we will set penalization strength λ to the optimal value λ^* disclosed for the LP, to simplify our inquiry. Simple Average will only be used at "One-Time" Estimation because its weights cannot be re-computed. Consequently, both approaches for the BTLP are applied in the Monte Carlo simulation as discussed before (subsection 4.3), with 1000 replications due to the large computation time of the methods. Finally, we apply the BTLP to the Eurozone inflation density forecasts of the SPF (subsection 4.4), including a translation to real interest rate using Fisher's equation. To quantitatively compare the performance of aggregation techniques, we perform a Diebold-Mariano (DM) t-test (Diebold & Mariano, 1995) for the LS and RS (see Equations 1 and 10 respectively) of these models, defined as:

$$t^{DM} = \sqrt{T}\frac{\bar{\mathrm{S}}_{LP} - \bar{\mathrm{S}}_{BTLP}}{\hat{\sigma}}, \tag{15}$$

where $\hat{\sigma} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\Big(\mathrm{S}(\tilde{p}_{LP,t}; y_t) - \mathrm{S}(\tilde{p}_{BTLP,t}; y_t)\Big)^2}$ and $\bar{\mathrm{S}}$ denotes the average score, so $\bar{\mathrm{S}}_{LP} = \frac{1}{T}\sum_{t=1}^{T}\mathrm{S}(\tilde{p}_{LP,t}; y_t)$. Under a null hypothesis of scores being equal (so same predictive ability), t^{DM} is asymptotically standard normally distributed (Gneiting & Katzfuss, 2014). Furthermore, to compare calibration between LP and BTLP we again use histograms of the PIT series. Next to this visual assessment, we formally check uniformity of the PIT series for which the Kolmogorov-Smirnov (KS) test is a straightforward, sensible option (Mitchell & Wallis, 2011) defined as:

$$D_T = \sup_y |F_T(y) - F(y)|, \tag{16}$$

where $F(y)$ denotes the theoretical CDF of the uniform distribution and $F_T(y)$ the empirical CDF given by:

$$F_T(y) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}_{[-\infty,y]}(y_t), \tag{17}$$

where $\mathbb{1}_{[-\infty,y]}(y_t)$ denotes the indicator function, so returning 1 if $y_t \leq y$ and 0 otherwise. If the test statistic D_T is large, we reject the null hypothesis of similar distributions.

5 Results

In the following sections, we will first present the results of the simulation for both linear aggregation and the BTLP (subsection 5.1 and 5.2). Second, we do the same for our empirical application to the Eurozone inflation density forecasts (subsection 5.3 and 5.4).

5.1 Results Monte Carlo Simulation

To evaluate the performance of the regularized mixture methods, we start by performing a Monte Carlo simulation with 10 000 replications (subsection 4.3). Outcomes for all methods are similar to the ones of Diebold et al. (2021), but of course slightly different due to the generation of random variables during the simulation. Below, in Table 3, the results are shown for DGP 1.

Table 3

Monte Carlo simulation results of all (regularized) methods for DGP 1

Method	LS	#Predictors	λ^*
Simple Average	-1.118	20.000	-
Simplex	-1.277	5.057	-
Simplex+Ridge	-1.118	20.000	10000.000
Simplex+Entropy	-1.118	20.000	20.000
Best 4-Average	-1.248	4.000	-
Best ≤4-Average	-1.331	3.287	-

LS denotes the average log score, #Predictors denotes the average number of used predictors, λ^* denotes the (ex-post) optimal penalty parameter.

Here, the superior performance of the Simple Average approach becomes clear, which is in accordance with the equal forecast signals featured in DGP 1. Essentially, the estimation error which arises at the computation of the Simplex weights, eventually causes a lower average log score compared to simply averaging the predictors, resembling the "equal weights puzzle" (Aastveit et al., 2018). This phenomenon becomes evident as well when determining the optimal lambda for Simplex+Ridge and Simplex+Entropy. Evaluating Figure 1 below, we observe the average log score to be monotonically increasing in the penalization strength (20 assessed values of λ) for Simplex+Ridge and Simplex+Entropy:

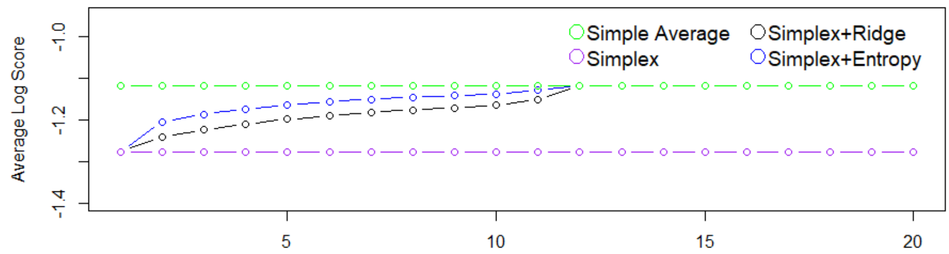


Figure 1: Average Log Score of regularization methods with respect to penalization strength, for DGP 1.

Therefore, the optimal penalization strength for both methods is located at the right bound of the searched intervals. Hence, as can be seen in Table 3, $\lambda^* = 10000.000$ for Simplex+Ridge and $\lambda^* = 20.000$ for Simplex+Entropy. Such, as penalization towards equal weights is so strong, the Simple Average approach is automatically imposed, reflected by the similar outcomes for these three methods all beating Best 4-Average and Best $\leq$4-Average. Although intuitively Best $\leq$4-Average might be expected to be the better of these two as there are more weighting possibilities, it performs worse than the Best 4-Average method and is beaten by all other approaches. Still, the average used predictors for Best $\leq$4-Average (3.287) is not that low compared to selecting exactly 4. When it comes to DGP 2, the (comparative) behaviour of the regularization methods is quite different from the one for DGP 1, as can be seen below in Table 4.

Table 4

Monte Carlo simulation results of all (regularized) methods for DGP 2

Method	LS	#Predictors	λ^*
Simple Average	-1.592	20.000	-
Simplex	-1.286	4.595	-
Simplex+Ridge	-1.168	8.692	15.000
Simplex+Entropy	-1.242	20.000	0.070
Best 4-Average	-1.236	4.000	-
Best $\leq$4-Average	-1.328	3.225	-

LS denotes the average log score, #Predictors denotes the average number of used predictors, λ^* denotes the (ex-post) optimal penalty parameter.

Now, the Simple Average approach performs the worst, which is due to the unequal forecast signals characterizing DGP 2. Such a feature causes the Simplex method to be more accurate as more importance is shifted to forecasters with high-quality signals. Naturally, estimation errors for Simplex weights still prevail, but eventually, this methods returns a higher average log score than for Simple Average. Furthermore, Figure 2 (presented below) displays an interesting pattern for the Simplex+Ridge and Simplex+Entropy performance with respect to penalization strength.

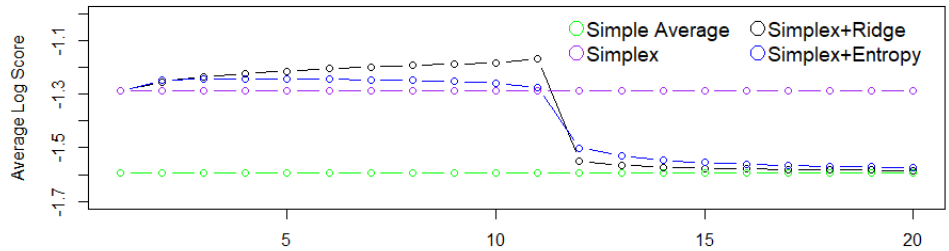


Figure 2: Average Log Score of regularization methods with respect to penalization strength, for DGP 2.

Starting at Simplex (no penalization), the average log score increases until it hits an optimum and decreases below Simplex, eventually coinciding with Simple Average (maximal penalization). Therefore, as can be seen in Table 4, $\lambda^* = 15.000$ for Simplex+Ridge and $\lambda^* = 0.070$ for Simplex+Entropy. For the interval of λ for which these two methods outperform Simplex, the strength of the penalization allows for shifting importance from bad to good performing forecasters but restraining weights to equality to evade estimation errors. Such, an efficient combined density forecast is constructed, reflected by the highest average log score for the Simplex+Ridge method displayed in Table 4. Besides, for this approach, 8.692 predictors are included on average compared to 4.595 for Simplex, revealing the regularization of Simplex+Ridge towards more equal weights. Furthermore, the Best 4-Average and Best $\leq$4-Average approaches now beat the Simple Average for DGP 2, where Best $\leq$4-Average is the least accurate of the two.

Concluding, the (comparative) behaviour of regularization methods depends on the forecast signals characterizing the DGP. While for equal disturbances the Simple Average is the most efficient due to the estimation error arising when weights are optimized (DGP 1), we observe unequal forecast signals to cause a relatively better performance for all other regularized mixture methods (DGP 2). In the latter case, Simplex+Ridge and Simplex+Entropy are preferred over Simplex for moderate penalization strength as they shrink weights to equality, reducing estimation error. For both DGP specifications, Best $\leq$4-Average is outperformed by Best 4-Average and selects a reasonable large amount of forecasters.

5.2 Results Monte Carlo Simulation Beta-Transformed Linear Pool

Moving on, we combine density forecasts using the BTLP and the regularized mixtures in the same simulation (now 1000 replications). The outcomes, including the average estimates of α and β are reported below in Table 5, both for DGP 1 and 2.

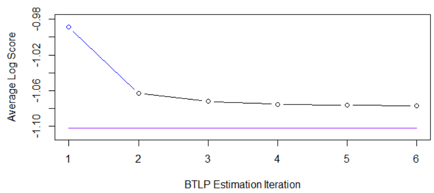
Table 5

Monte Carlo simulation results of all (regularized) methods applied to both the LP and BTLP, DGP 1 and 2

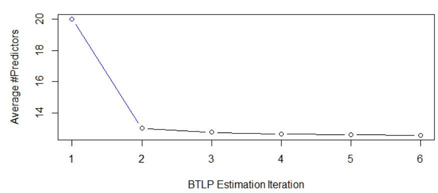
Method	Aggregation technique, estimation method	DGP 1 LS	DGP 1 #Predictors	DGP 1 α^*	DGP 1 β^*	DGP 2 LS	DGP 2 #Predictors	DGP 2 α^*	DGP 2 β^*
Simple Average	Linear Pool	-1.102	20.000	-	-	-1.605	20.000	-	-
	BTLP, "One Time"	-0.989	20.000	1.377	1.493	-1.425	20.000	1.411	1.678
Simplex	Linear Pool	-1.251	5.063	-	-	-1.347	4.652	-	-
	BTLP, "One Time"	-1.252	5.063	1.309	1.401	-1.372	4.652	1.273	1.334
	BTLP, Alternating	-1.233	6.505	1.313	1.545	-1.299	5.535	1.294	1.443
Simplex+Ridge	Linear Pool	-1.102	20.000	-	-	-1.199	8.744	-	-
	BTLP, "One Time"	-0.989	20.000	1.377	1.493	-1.144	8.744	1.340	1.449
	BTLP, Alternating	-1.077	12.595	1.351	1.607	-1.160	9.085	1.337	1.517
Simplex+Entropy	Linear Pool	-1.102	20.000	-	-	-1.287	20.000	-	-
	BTLP, "One Time"	-0.989	20.000	1.377	1.494	-1.243	20.000	1.325	1.439
	BTLP, Alternating	-1.097	20.000	1.345	1.584	-1.228	20.000	1.324	1.528
Best 4-Average	Linear Pool	-1.193	4.000	-	-	-1.296	4.000	-	-
	BTLP, "One Time"	-1.174	4.000	1.324	1.417	-1.300	4.000	1.286	1.356
	BTLP, Alternating	-1.214	4.000	1.330	1.495	-1.299	4.000	1.301	1.400
Best ≤4-Average	Linear Pool	-1.286	3.261	-	-	-1.399	3.241	-	-
	BTLP, "One Time"	-1.299	3.261	1.294	1.357	-1.449	3.241	1.256	1.297
	BTLP, Alternating	-1.317	3.545	1.316	1.456	-1.402	3.450	1.284	1.358

LS denotes the average log score, #Predictors denotes the average number of used predictors, α^* denotes average optimal α, β^* denotes average optimal β.

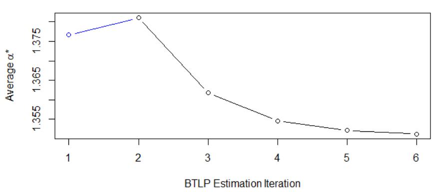
For "One-Time" Estimation, we observe the log score for the BTLP increases compared to the LP for Simple Average, Simplex+Ridge, Simplex+Entropy, while it decreases for Simplex (but only slightly) and Best ≤4-Average for both DGP types. Only for Best 4-Average the log score increases for DGP 1, while not for DGP 2. Naturally, as we used the optimal penalization strength (λ) for Simplex+Ridge and Simplex+Entropy, which is the largest possible for both methods for DGP 1, the results of Simple Average and these two methods are all very similar, contrary to DGP 2. For "One-Time" Estimation, α^* ranges from about 1.2 to 1.4 and β^* from 1.3 to 1.7. As we do not change the weights for this estimation method, the average number of used predictors is equal to the ones for linear aggregation. In addition, this shows the change of the log score at "One-Time" Estimation compared to the LP is solely caused by the beta-transformation, while for Alternating Estimation, an effect caused by the re-estimation of weights appears. To better understand the course along with the Alternating Estimation, we present graphs of the simulation outcomes for Simplex+Ridge, DGP 1 with respect to estimation iteration below in Figure 3 (see Appendix D6 for the other regularized methods and DGP types).



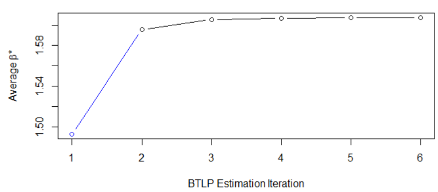
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors



(c) Average α^*



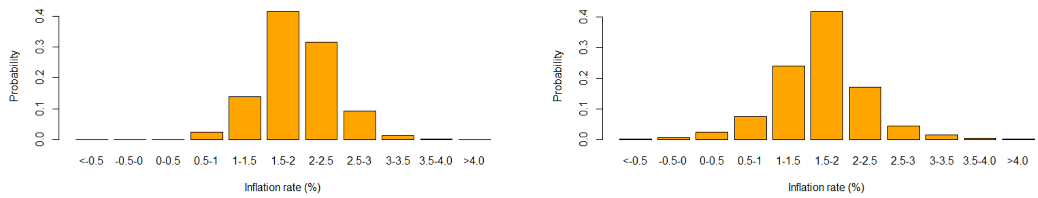
(d) Average β^*

Figure 3: Simulation results for the BTLP with respect to estimation iteration, for Simplex+Ridge, DGP 1. First "One-Time" Estimation in blue.

Here, we see some convergence as the log score gradually decreases over iterations, but always staying above the log score for linear aggregation. However, as the average number of predictors decreases too (and we know fewer predictors implies a lower log score, in general), the effect of the BTLP itself reflected by the graphs for α^* and β^* is hard to extract. For this estimation technique, we observe in Table 5 improved log scores only for Simplex compared to "One-Time" Estimation for DGP 1, but as well for Simplex+Entropy, Best 4-Average and Best $\leq$4-Average for DGP 2 (but the latter two not beating linear aggregation). At the other regularized methods, the log score drops. The relatively good performance of Simplex (and Simplex+Entropy with modest penalization strength at DGP 2) during the Alternating Estimation might be explained by its ability to estimate weights most "freely" compared to the other regularized methods. Furthermore, all α^* and β^* are larger than for "One-Time" Estimation, implying we move further away from linear aggregation (for which $\alpha = \beta = 1$). Immediately, this might explain the log score behaviour for the Alternating Estimation, as this probably depends on whether we correctly estimate weights, α and β, moving towards the "right direction". This is what happens on average for Simplex as log score increases, but we "move into" the wrong model for the other methods due to their regularization. Hence, in general, we might prefer the "One-Time" Estimation. These are informative theoretical results, but to assess the practical performance of the regularized mixtures we now turn to our empirical application, first with linear aggregation as in Diebold et al. (2021), second with the BTLP.

5.3 Results Empirical Application

Starting with our empirical application to Eurozone inflation, we first resolve the data issues discussed in subsection 4.4, leaving 18 forecasters. As an initial assessment of the produced dataset after these refinements, we present the average of the predictive densities (not including the uniform forecaster) for 2004Q4 and 2018Q4 below in Figure 4:



(a) 2004Q4 **(b)** 2018Q4

Figure 4: Average of predictive densities for Eurozone inflation, 2004Q4 and 2018Q4.

These histograms match the average predictive densities presented by Diebold et al. (2021) for these periods and reflect the shift towards fatter tails for later surveys. Moving on, we implement the regularization methods using the rolling window with length of 20 quarters (see subsection 4.4), replicating Diebold et al. (2021). Below in Table 6, the results are presented:

Table 6

Results of all (regularized) methods for Eurozone inflation

Method	LS	#Predictors	λ^*
Simple Average	-2.002	19.000	-
Simplex	-1.996	3.840	-
Simplex+Ridge	-1.944	9.053	10.000
Simplex+Entropy	-1.961	19.000	0.130
Best 4-Average	-1.976	4.000	-
Best $\leq$4-Average	-2.001	2.640	-
Individual Forecasters	**LS**	**#Predictors**	**λ^***
Best	-2.012	1.000	-
90%	-2.091	1.000	-
70%	-2.203	1.000	-
Median	-2.220	1.000	-
Worst	-2.700	1.000	-
Uniform	-2.398	1.000	-

LS denotes the average log score, #Predictors denotes the average number of used predictors, λ^* denotes the (ex-post) optimal penalty parameter.

Evaluating Table 6, the best-performing method is Simplex+Ridge, while Simple Average shows the lowest log score. Comparing the methods with the individual forecasters, the best predictor is still beaten by all mixture methods. Besides, Simplex+Ridge performs 12.4% better compared to the median forecaster, reflecting the large gains which can be achieved when implementing this method. Remarkably, the uniform forecaster has a log score of -2.398, outperforming three of the eight-teen survey respondents. Below in Figure 5, we see the pattern of (regularized) mixture performance is similar to the one encountered in the simulation for DGP 2.

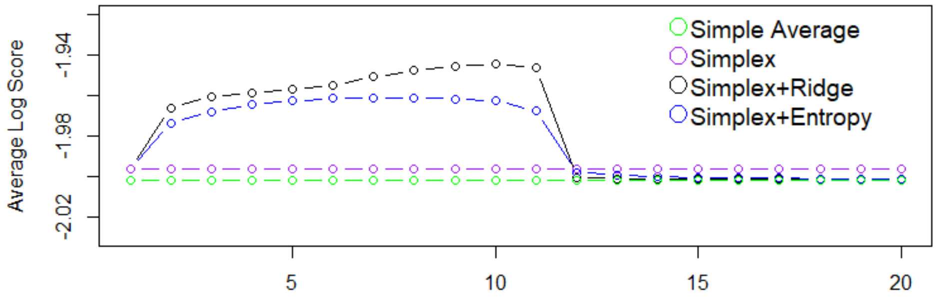


Figure 5: Average Log Score of regularization methods with respect to penalization strength, for Eurozone inflation.

Naturally, the forecasting signals of the predictors in the SPF survey are not all of the same quality, increasing the relative performance of Simplex compared to Simple Average (as in DGP 2). Nonetheless, the average log scores of Simplex and Simple Average lie very close, implying the estimation error of the Simplex weights drastically reduces its performance. Furthermore, Best 4-Average performs better than Best ≤4-Average, in line with our simulation. Nonetheless, Best ≤4-Average now selects fewer predictors, which might be an explanation for its relatively low average log score. Below, in Figure 6, we show the resulting inflation combined density forecasts for the Simple Average and Simplex+Ridge method.

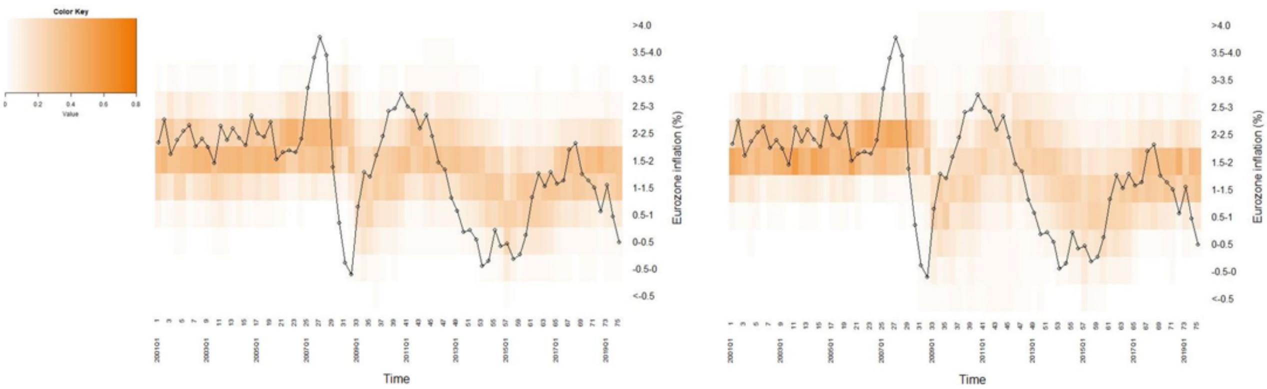


Figure 6: Simple Average (left) and Simplex+Ridge (right) combined density forecasts for (superimposed) Eurozone inflation 2001Q1 - 2019Q3, displayed as heatmap.

Here, the differences between the Simple Average and a regularized mixture method, Simplex+Ridge, become clear. For Simple Average, the level of dispersion is about the same over the entire sample period. However, for Simplex+Ridge, densities are relatively dispersed for the first observations ranging from 2001Q1 - 2006Q4 and are shifted a bit upwards compared to

Simple Average. After that, during the quarters of the Great Recession 2007Q1-2008Q4, realized inflation shows a large spike after which it drops and builds up again. Such, a shift begins towards fatter tails of the combined densities due to the regularization effects of Simplex+Ridge, which can be seen in Figure 6 by the covering of a larger plane in somewhat lighter orange for 2009Q1 - 2019Q3. Furthermore, the distinction between Simple Average and Simplex+Ridge is revealed by a heatmap of the differences between these combined densities below in Figure 7.

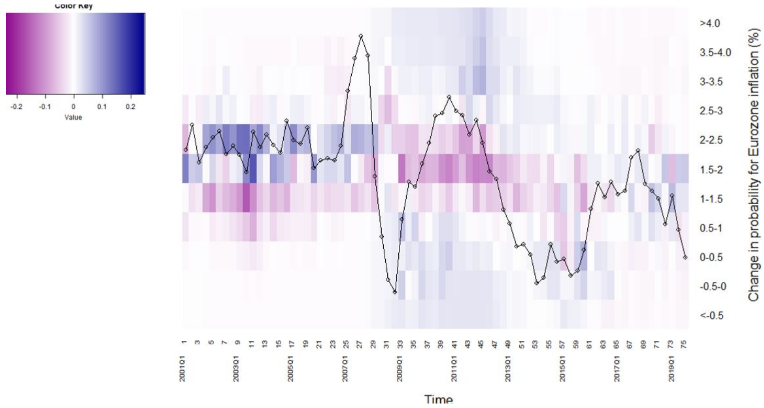
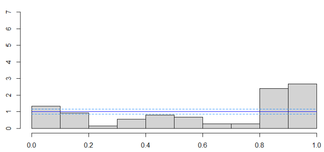
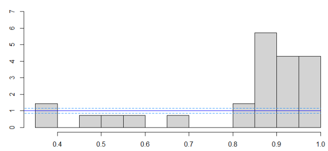


Figure 7: Heat map of differences between Simplex+Ridge and Simple Average combined density forecasts for (superimposed) Eurozone inflation 2001Q1-2019Q3 (Simplex+Ridge minus Simple Average).

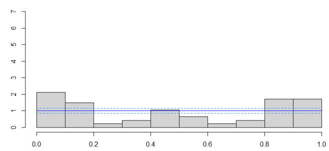
Again, the regularization effects after the Great Recession come forward as light blue fields appear at the upper and lower parts of the heatmap around 2009Q1-2015Q1, indicating Simplex+Ridge moves more probability mass towards the tails compared to Simple Average. Moreover, at the start of the period, around 2001Q1-2005Q1, it becomes clear Simplex+Ridge regularization moves the distribution slightly upwards. Still, many observations in this period fall in the right-tail of the combined densities as revealed by the histograms of the PIT series (which should be *iid* uniform under ideal forecasting) for Simple Average and Simplex+Ridge displayed below in Figure 8 and 9.



(a) Full sample

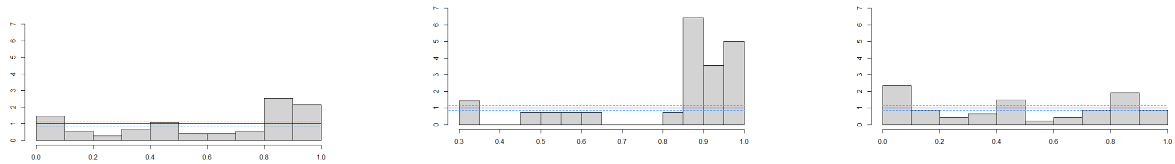


(b) First subsample, 2001Q1 - 2007Q4



(c) Second subsample, 2008Q1 - 2019Q3

Figure 8: Histograms of the PIT series for the Simple Average method, Eurozone inflation 2001Q1-2019Q3. 95% confidence intervals are dotted in light blue.



(a) Full sample

(b) First subsample, 2001Q1 - 2007Q4

(c) Second subsample, 2008Q1 - 2019Q3

Figure 9: Histograms of the PIT series for the Simplex+Ridge method, Eurozone inflation 2001Q1-2019Q3. 95% confidence intervals are dotted in light blue.

Although Simplex+Ridge implements a correction upwards, the histograms for both methods show to be highly biased for the first subsample (2001Q1-2007Q4). However, the Simplex+Ridge histogram for the second subsample (2008Q1-2019Q3) is more uniform compared to Simple Average, reflecting the regularization effects toward fatter tails after the Great Recession. Hence, for the entire sample, Simplex+Ridge eventually causes improved calibration. In Appendix D1, D2 and D3 we respectively show the combined density forecasts, heatmaps of differences between Simple Average, and PIT histograms for the other regularized methods. These all show results which are comparable with Simplex+Ridge, such as the regularization effects after the Great Recession. Moreover, in line with Diebold et al. (2021), the Simplex and Best ≤4-Average methods display very similar results (also the difference in log score found in Table 6 is quite small), which is confirmed by a heatmap of differences between the combined densities presented below in Figure 10.

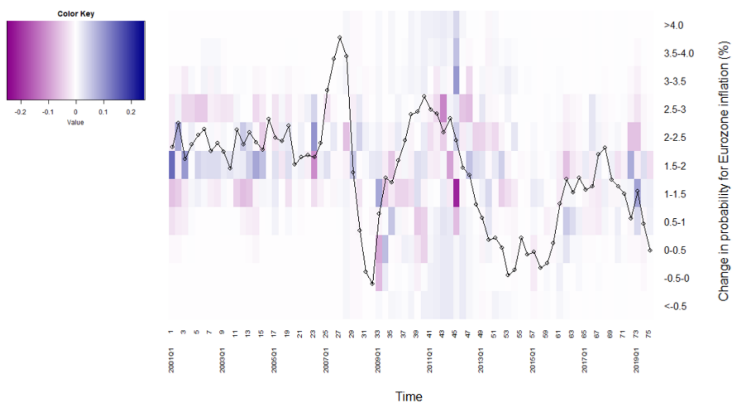


Figure 10: Heat map of differences between Simplex and Best ≤4-Average combined density forecasts (Simplex minus Best ≤4-Average). Realized inflation is superimposed.

Finally, we transform the regularized mixtures for Eurozone inflation to real interest rate density forecasts by using Fisher's transformation. Below in Figure 11, we show this combined density forecast for Simplex+Ridge.

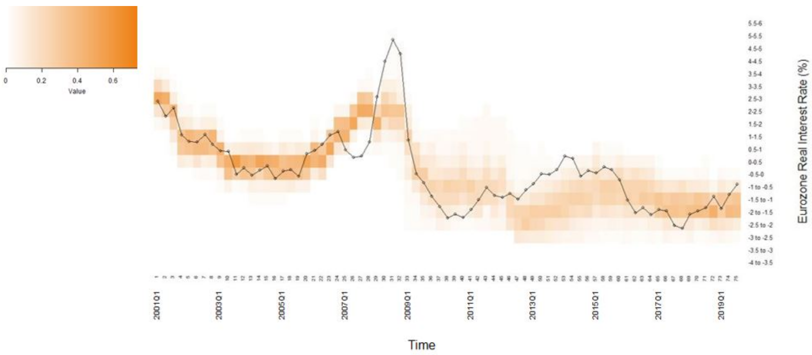


Figure 11: Simplex+Ridge combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1 - 2019Q3 (75 quarters), displayed as heatmap.

By construction, this is a location shift (of the interest rate at t for $t = 1, ..., T$) and mirror of the inflation density forecasts. Therefore, we observe the same patterns of dispersion in Figure 11 as in Figure 6 and do not report average log scores and PIT histograms (see subsection 4.4). Still, we show a heatmap of the differences between Simplex+Ridge and Simple Average below in Figure 42:

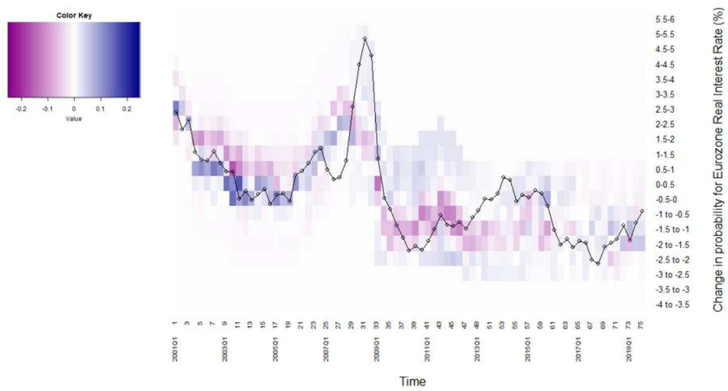


Figure 12: Heat map of differences between Simplex+Ridge and Simple Average combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1-2019Q3 (Simplex+Ridge minus Simple Average).

Here, the discussed regularization effects of Simplex+Ridge compared to Simple Average remain to be visible.

All findings presented above on regularized combining macroeconomic density forecasts are similar to the ones of Diebold et al. (2021), although results such as the log scores of Table 6 do not match exactly, which can be explained by several factors. First, the inclusion of the uniform forecasters leads to increased performance of the Simple Average method, justifying its closer proximity to the regularized mixtures. Second, as Diebold et al. (2021) do not precisely specify their moving-window implementation, the most sensible choice would be to follow Diebold and Shin (2019), but possibly their approaches differ. Finally, the used data (SPF, or realizations on inflation and real interest rate) might have been revised in the meantime. Nonetheless, the comparative behaviour of regularized mixtures, regularization effects after the Great Recession and similarity of Simplex and Best $\leq$4-Average are all in line with Diebold et al. (2021).

5.4 Results Empirical Application Beta-Transformed Linear Pool

Extending to Diebold et al. (2021), we apply the BTLP to Eurozone inflation density forecasts, using the same regularized mixtures as discussed. Below in Table 7, we present the results.

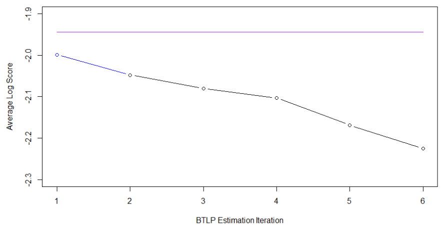
Table 7

Empirical results of all (regularized) methods applied to both the LP and BTLP, Eurozone inflation

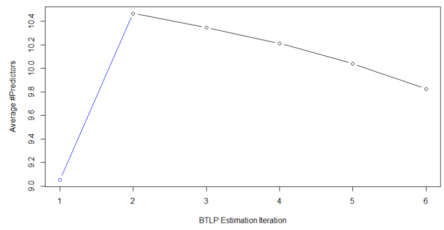
Method	Aggregation technique, estimation method	LS	t-test with LP, for LS	RS	t-test with LP, for RS	#Predictors	α^*	β^*
Simple Average	Linear Pool	-2.002	-	1.026	-	19.000	-	-
	BTLP, "One Time"	-2.050	0.421	1.037	-0.197	19.000	2.188	0.985
Simplex	Linear Pool	-1.996	-	1.000	-	3.840	-	-
	BTLP, "One Time"	-2.040	0.429	1.021	-0.429	3.840	1.273	1.021
	BTLP, Alternating	-2.430	1.134	1.070	-1.048	3.360	3.129	0.860
Simplex+Ridge	Linear Pool	-1.944	-	0.987	-	9.053	-	-
	BTLP, "One Time"	-1.999	0.627	1.023	-0.743	9.053	1.923	1.028
	BTLP, Alternating	-2.225	1.066	1.058	-1.190	9.827	3.130	0.940
Simplex+Entropy	Linear Pool	-1.961	-	0.996	-	19.000	-	-
	BTLP, "One Time"	-1.999	0.397	1.017	-0.414	19.000	2.073	1.041
	BTLP, Alternating	-2.237	1.027	1.057	-1.027	19.000	2.813	0.994
Best 4-Average	Linear Pool	-1.976	-	1.001	-	4.000	-	-
	BTLP, "One Time"	-2.011	0.388	1.036	-0.683	4.000	1.903	1.009
	BTLP, Alternating	-2.091	0.666	1.084	-1.234	4.000	2.971	0.893
Best $\leq$4-Average	Linear Pool	-2.001	-	1.001	-	2.640	-	-
	BTLP, "One Time"	-2.048	0.479	1.038	-0.757	2.640	1.866	0.989
	BTLP, Alternating	-2.672	0.986	1.121	-1.564	2.427	3.043	0.828

LS denotes the average log score, RS denotes the average ranked score, #Predictors denotes the average number of used predictors, α^* denotes average optimal α, β^* denotes average optimal β.

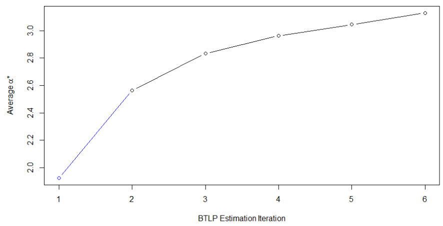
Here, it becomes clear the log score decreases for all methods decreases when moving from the LP to the BTLP. Although for "One-Time" Estimation this is only a small difference, the Alternating Estimation approach performs quite bad, even for Simplex. The same patterns are observed for RS (which should be low for good predictive performance), for which the score is always higher when implementing the BTLP. Nonetheless, DM t-tests for both the LS and RS disclose no significant differences in score between LP and BTLP. Still, these results seem to be a disappointment when relating to our simulation, for which multiple methods showed increased performance when implementing the BTLP. This difference could be explained by the simulation set-up, which is designed to evaluate regularized mixture methods and not necessarily the aggregation technique. However, a key insight of the simulation remains relevant, namely the detrimental effects on performance when moving towards the wrong model during the Alternating Estimation. Below in Figure 13, we present the outcomes for Simplex+Ridge with respect to estimation iteration (for the other regularized methods see Appendix D7).



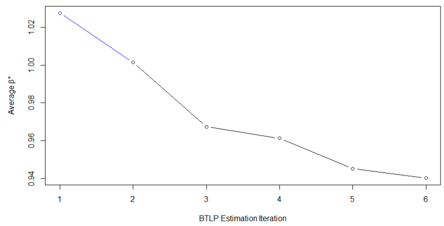
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors



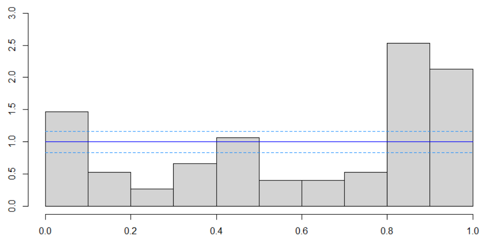
(c) Average α^*



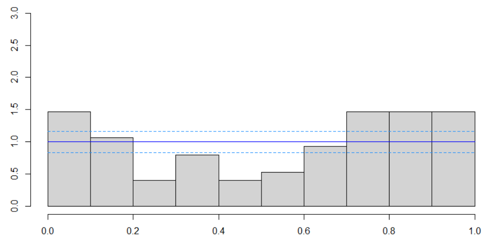
(d) Average β^*

Figure 13: Empirical results for the BTLP with respect to estimation iteration, for Simplex+Ridge, Eurozone inflation. First "One-Time" Estimation in blue.

These graphs show the rapid decrease of the log score over the iteration, in which the number of predictors might play a role as it decreases too, indicating we might end up in the wrong direction. Therefore, it would be sensible to stick to "One-Time" Estimation, although it still shows a lower log score compared to the LP. However, a major advantage of the BTLP ("One-Time" Estimation) appears when considering calibration of the combined density forecasts as assessed through the PIT histograms. Below in Figure 14, we display this for the Simplex+Ridge method (showing the highest log score together with Simplex+Entropy for "One-Time" Estimation) for both the LP and BTLP.



(a) Linear Pool



(b) Beta-Transformed Linear Pool

Figure 14: PIT histograms Simplex+Ridge for Eurozone inflation, using the LP and BTLP ("One-Time" Estimation). 95% confidence intervals are dotted in light blue.

As the PIT series should be *iid* uniformly distributed for ideal forecasting, these histograms show a large improvement when moving to the BTLP, which is the case for all (regularized) methods as can be seen in Appendix D8. While the PIT histogram is severely biased for linear aggregation, with many observations falling in the extreme right tails of the forecasted distributions, the BTLP almost entirely solves this. As can be seen in Figure 14b, the bins are quite close to each other with no large spikes, only showing slight under-dispersion. In addition to this visual assessment, the improvement of the BTLP ("One-Time" Estimation) is disclosed quantitatively by the KS tests for uniformity of the PIT series displayed below in Table 8.

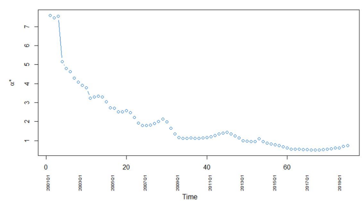
Table 8

Kolmogorov-Smirnov tests for uniformity of the PIT series for all (regularized) methods, under the LP and BTLP ("One-Time" Estimation)

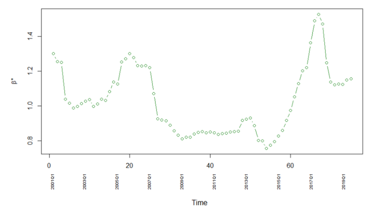
Method	KS test, LP	KS test, BTLP
Simple Average	0.311**	0.135
Simplex	0.252**	0.132
Simplex+Ridge	0.271**	0.161*
Simplex+Entropy	0.274**	0.168*
Best 4-Average	0.255**	0.152
Best ≤4-Average	0.272**	0.130

* = significant at 5%-level, ** = significant at 1%-level

Here, we see uniformity of the PIT series to be rejected for all regularized mixtures under the LP. However, for the BTLP the KS test only rejects for Simplex+Ridge and Simplex+Entropy at the 5% significance level and not for the other methods. To further investigate this improvement brought about by the BTLP, we consider the estimates of α^* and β^* over our sample period, displayed below in Figure 15 for Simplex+Ridge.



(a) α^*



(b) β^*

Figure 15: "One-time" estimates of α^* and β^* for the BTLP over time, for Simplex+Ridge, Eurozone inflation.

To understand the patterns in these graphs (also in Appendix D9 for the other regularized methods) and the flexible dispersion it brings about, we display CDF's of the Beta distribution used in the BTLP (with Simplex+Ridge) under α^* and β^* for 2008Q2 (a crisis year), 2019Q3 (the final observation), on average and for linear aggregation below in Figure 16.

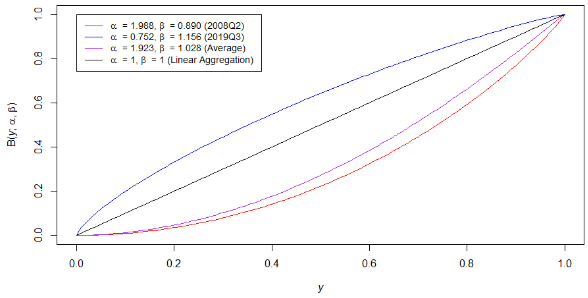


Figure 16: CDF's of the Beta distribution exploited for the BTLP and Simplex+Ridge method for Eurozone inflation (2008Q2, 2019Q3, on average and for linear aggregation).

Recalling Equation 11, the BTLP is implemented by evaluating the CDF of the linear combination in the CDF of the beta distribution as displayed above in Figure 16. For linear aggregation, naturally, $\alpha = \beta = 1$ such that we have a 45° line and the same CDF is returned. For 2008Q2 (in red), we see a parabolic shape completely below this line, indicating BTLP corrects the CDF downwards. Such, while for linear aggregation the observed inflation would fall in a CDF bin with loading 0.923, so this is the PIT for 2008Q2, BTLP corrects this downwards to 0.875. On the other hand, for 2019Q3 (in blue), the Beta CDF lies above the 45° line, implying the BTLP corrects the CDF upwards, adjusting the PIT value from 0.038 to 0.0956. Hence, the graphs of Figure 15 reflect these corrections α^* and β^* jointly bring about over time. So, for instance, β^* drops during the Great Recession quarters (2007Q1-2008Q4) to anticipate for the high spike in inflation level. Thus, it would be fruitful to implement the BTLP ("One-Time Estimation"), at the cost of a slightly and insignificantly lower average log score, for its large calibration advantages. Therefore, we present the combined density forecast for Simplex+Ridge using the BTLP below in Figure 17 (which we do for the other regularized methods in Appendix D10).

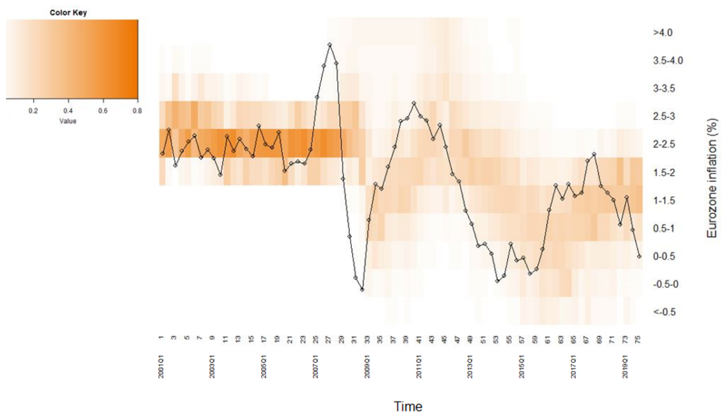


Figure 17: Simplex+Ridge combined density forecasts for (superimposed) Eurozone inflation 2001Q1 - 2019Q3 for the BTLP ("One-Time" Estimation), displayed as heatmap.

Comparing this graph with the linearly combined density forecasts for Simplex+Ridge displayed in Figure 6, we see a similar pattern, especially when it comes to the regularization effects after the Great Recession. Still, there are some differences too as expressed by the heatmap below in Figure 18 (similar graphs in Appendix D11 for other regularized mixtures).

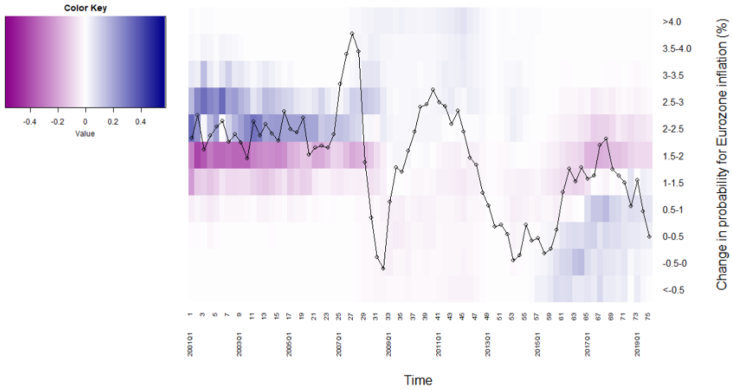


Figure 18: Heat map of differences between the BTLP and LP for Simplex+Ridge combined density forecasts for (superimposed) Eurozone inflation 2001Q1-2019Q3 (BTLP minus LP).

This shows the BTLP combined density forecasts to be located somewhat more upwards compared to the LP for 2001Q1 - 2008Q4. After that, for 2009Q1-2014Q4, the differences are smaller, but the BTLP clearly moves probability mass to the right tails of the predictive densities to account for potentially higher inflation outcomes. From 2015Q1 till the end of the sample, this pattern reverses and bins at the left-hand side of the distribution are assigned higher probabilities compared to the LP, reacting to the large decline in realized inflation starting in 2012Q1.

Consequently, we again compare the Best ≤4-Average and Simplex method under the BTLP by considering the heat map of differences between these densities presented below in Figure 19.

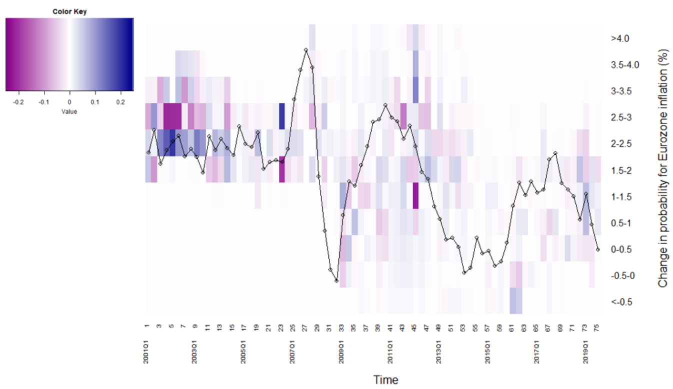


Figure 19: Heat map of differences between Simplex and Best ≤4-Average combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1-2019Q3 for the BTLP (Simplex minus Best ≤4-Average).

Although there are some differences for 2001Q1-2004Q1 with Best ≤4-Average assigning somewhat higher probabilities to right-tail bins compared to Simplex, for the rest of the sample period the distributions lie very close to each other, similar to linear aggregation. Finally, we transform to real interest rate predictive densities using Fisher's equation for the BTLP, of which the Simplex+Ridge combined density forecasts are presented below in Figure 20 (similar graphs in Appendix D12 for other regularized mixtures).

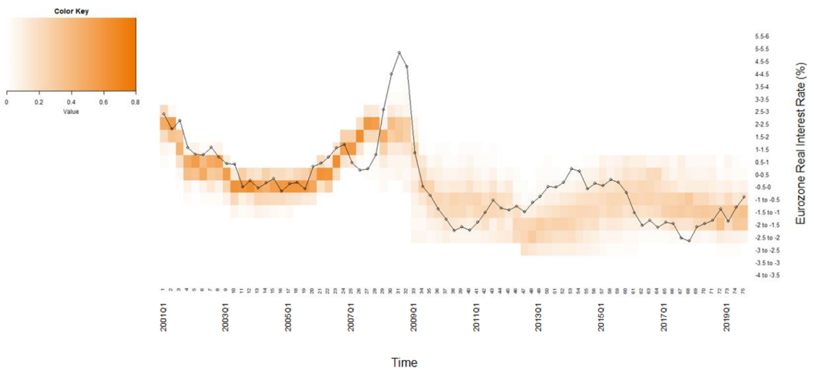


Figure 20: Simplex+Ridge combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1 - 2019Q3 for the BTLP ("One-Time" Estimation), displayed as heatmap.

Again, this is a location shift and mirror of the graph of Figure 17. Therefore, the conclusions stemming from Figure 18 still hold, but in opposite direction: now, the BTLP moves the density forecasts more downwards compared to linear aggregation for 2001Q1-2008Q4, for instance.

6 Conclusion & Discussion

In this paper, we combine Eurozone macroeconomic variables using the (beta-transformed) linear pool and various regularized mixtures as proposed by Diebold et al. (2021). Following their research, we apply these methods in a Monte Carlo simulation, using two DGP types to distinguish between forecasting signal quality. Such, we find Simplex+Ridge with modest penalization strength to perform best for unequal disturbances, which we too encounter when combining Eurozone inflation density forecasts retrieved from the SPF, outperforming the best individual predictors. Moreover, distribution tails grow fatter after the Great Recession due to regularization effects and Simplex and Best $\leq$4-Average perform identically, in line with Diebold et al. (2021). Furthermore, calibration evaluated through PIT histograms is rather poor.

However, we greatly improve this when implementing the BTLP proposed by Gneiting and Ranjan (2013), allowing for the flexible dispersion lacked by linear mixtures, in particular in this macroeconomic application of combining relatively few predictors (Aastveit et al., 2018). Introducing two methods to assess the α and β used for this class, "One-Time" and Alternating Estimation, we start by applying this non-linear aggregation technique to the simulation. Here, the potential of the BTLP to ameliorate forecast combinations becomes clear as the average log score increases for many methods. Nonetheless, as an Alternating Estimation could move the model in the wrong direction, "One-Time" Estimation might be preferred, which is confirmed by the application to Eurozone inflation. Again, we encounter the regularization effects and similarity between Simplex and Best $\leq$4-Average as for linear aggregation, while the BTLP is able to more flexibly move probability mass to extremer bins. Although implementing the BTLP ("One-Time" Estimation) does lead to a slight and insignificant decline in average log score, PIT histograms become greatly smoother, showing improved calibration due to the flexible dispersion imposed by the BTLP, confirmed significantly by KS tests. Hence, we recommend using this class to aggregate Eurozone inflation density forecasts in combination with Simplex+Ridge.

Finally, we transform to real interest rate density forecasts using Fisher's equation. As this comes down to a simple location shift and mirror of the inflation distribution, all conclusions regarding the regularized mixtures (comparative behaviour, regularization effects, calibration) still hold, also for the BTLP. For the "Neo-Fisher" effect (Uribe, 2018), we even show simpler conversion should this phenomenon become more eminent.

Nonetheless, literature shows the Fisher equation to not rigorously hold in practice, especially for the long term in the Eurozone (Koustas & Serletis, 1999). Thus, it would be more accurate to use the transformation only in applications where the effect is properly identified (short vs long run, type of interest rate, etc.) or to construct density forecasts for real interest rate directly. Still, the latter approach would mean another survey has to be conducted, while a major advantage of the Fisher transformation is the straightforward recycling of inflation predictions.

Moreover, there are some methodological points of discussion. First, the optimal ex-post penalization strength λ is used for Simplex+Ridge and Simplex+Entropy, which would be infeasible in real time. Hence, we might assess the optimal λ together with the weights (and α and β for the BTLP) during the moving window estimation. This brings attention to another matter for future research, namely the length of the window. While Diebold et al. (2021) and this paper use 20 quarters, shorter or longer periods (10, 30, 40 quarters) might be more appropriate to respectively relate predictions closer to preceding observations or reduce estimation errors.

Furthermore, when it comes to the BTLP, more research is needed on the effect of wrongly estimating weights, α and β on the log score, in specific for its behaviour along with the iterations for Alternating Estimation. A sensible approach to start with is to design a simulation study for the BTLP to compare between using estimated and true parameters.

Finally, although the BTLP is a well-performing combination method judged on the literature and our results, the other non-linear techniques listed in Table 1 might be considered as well to truly reveal which aggregation type is most accurate. Furthermore, as this research specifically addresses Eurozone inflation, linear pooling or other non-linear methods might be more accurate for different applications of regularized mixtures.

References

Aastveit, K., Van Dijk, H., Mitchel, J., & Ravazzolo, F. (2018). The Evolution of Forecast Density Combinations in Economics. *Tinbergen Institute Discussion Papers*.

Bassetti, F., Casarin, R., & Ravazzolo, F. (2015). Bayesian nonparametric calibration and combination of predictive distributions. *Journal of the American Statistical Association*, *113*.

Bates, J., & Granger, C. (1969). The combination of forecasts. *Operational Research Society*, *20*(4), 451-468.

Bayer, S. (2017). Combining value-at-risk forecasts using penalized quantile regressions. *SSRN Electronic Journal*.

Booth, G., & Ciner, C. (2001). The relationship between nominal interest rates and inflation: International evidence. *Journal of Multinational Financial Management*, *11*, 269-280.

Bröcker, J., & Smith, L. (2007). Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecasting*, *22*(2), 382 - 388.

Clark, T., & McCracken, M. (2010). Averaging forecasts from vars with uncertain instabilities. *Journal of Applied Econometrics*, *25*(1), 5-29.

Cumings-Menon, R., & Shin, M. (2020). Probability forecast combination via entropy regularized wasserstein distance. *Entropy*, *22*, 929.

Czado, C., Gneiting, T., & Held, L. (2009). Predictive model assessment for count data. *Biometrics*, *65*(4), 1254-1261.

Diebold, F. (2000). *Elements of forecasting.* CENGAGE LEARNING.

Diebold, F., Gunther, T., & Tay, A. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*(4), 863-883.

Diebold, F., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business Economic Statistics*, *13*(3), 253-63.

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.

Diebold, F., Shin, M., & Zhang, B. (2021). On the aggregation of probability assessments: Regularized mixtures of predictive densities for eurozone inflation and real interest rates. *PIER Working Paper No. 21-002*.

Fama, E. F. (1975). Short-term interest rates as predictors of inflation. *The American Economic Review*, *65*(3), 269-282.

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

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

Glahn, B., Peroutka, M., Wiedenfeld, J., Wagner, J., Zylstra, G., Schuknecht, B., & Jackson, B. (2009). Mos uncertainty estimates in an ensemble framework. *Monthly Weather Review*, *137*(1), 246 - 268.

Gneiting, T., & Katzfuss, M. (2014). Probabilistic Forecasting. *The Annual Review of Statistics and Its Application*, *1*, 125-151.

Gneiting, T., & Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359-378.

Gneiting, T., & Ranjan, R. (2013). Combining predictive distributions. *Electronic Journal of Statistics*, *7*, 1747-1782.

Granger, C. (1989). Invited review combining forecasts-twenty years later. *Journal of Forecasting*, *8*, 167-173.

Hall, S., & Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, *23*, 1-13.

Jaffe, J., & Mandelker, G. (1976). The "fisher effect" for risky assets: An empirical investigation. *Journal of Finance*, *31*(2), 447-58.

Koustas, Z., & Serletis, A. (1999). On the fisher effect. *Journal of Monetary Economics*, *44*(1), 105-130.

Mishkin, F. (1991). *Is the fisher effect for real? a reexamination of the relationship between inflation and interest rates* (Working Paper No. 3632). National Bureau of Economic Research.

Mitchell, J., & Wallis, K. (2011). Evaluating density forecasts: forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics*, *26*(6), 1023-1040.

Raftery, A., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*.

Timmermann, A. (2006). Forecast Combinations. In G. Elliot & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 1, p. 135-196). Elsevier Science and Technology.

Uribe, M. (2018). *The neo-fisher effect: Econometric evidence from empirical and optimizing models* (Working Paper No. 25089). National Bureau of Economic Research.

Varadhan, R., & Gilbert, P. (2009). BB: An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. Retrieved from `http://www.jstatsoft.org/v32/i04/` (R package version 2019.10.1)

Wallis, K. (2005). Combining Density and Interval Forecasts: A Modest Proposal. *Oxford Bulletin of Economics and Statistics*, *67*(1), 983-994.

Weigel, A., Liniger, M., & Appenzeller, C. (2007). The discrete brier and ranked probability skill scores. *Monthly Weather Review*, *135*(1), 118 - 124.

Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average bayesian predictive distributions (with discussion). *Bayesian Analysis*, *13*(3), 917-1003.

Appendix A Overview Forecasting Literature

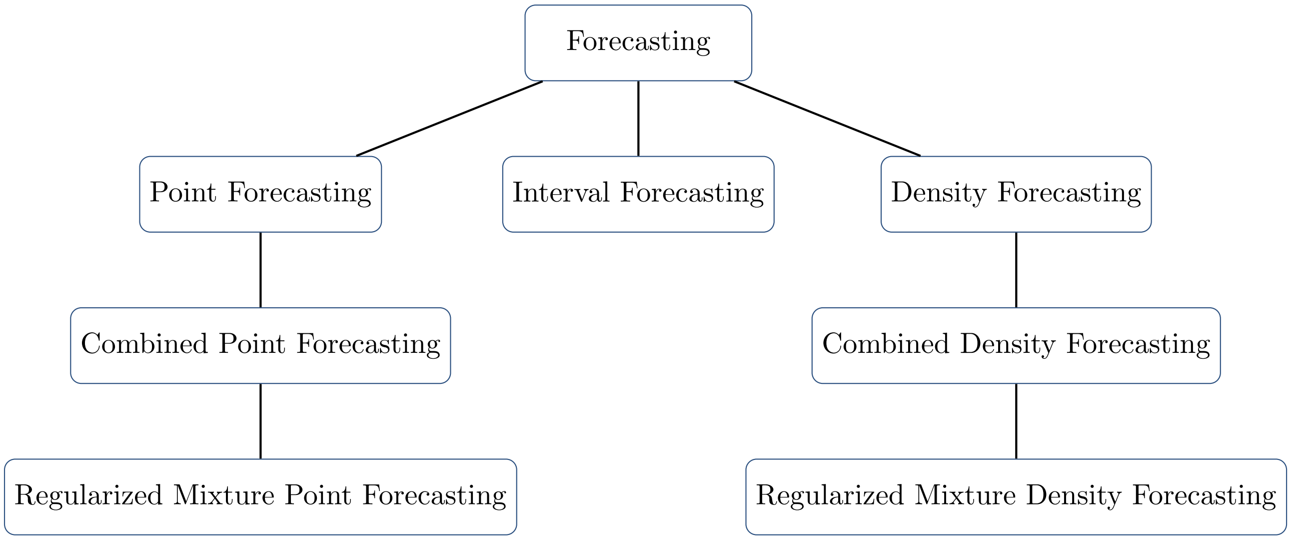


Figure 21: Schematic overview of the discussed literature on forecasting.

Appendix B Proofs Fisher Transformation

We use Fisher's equation $r_t = i_t - y_t$ for $t = 1, ..., T$ to transform inflation density forecast to real interest density forecasts by implementing a location shift and mirroring of the "original" inflation density forecasts, such that $p_t^r(\cdot) = i_t - p_t^\pi(\cdot)$ for $t = 1, ..., T$. Consequently, we prove the following two statements:

1. **The Log Score (LS) is invariant to the Fisher transformation.**

   We have to prove that $\log(p_t^r(r_t)) = \log(p_t^\pi(y_t))$ for $t = 1, ..., T$. Proof:

$$
\begin{aligned}
\log(p_t^r(r_t)) &= \log(i_t - p_t^\pi(r_t)) && \text{Fisher transformation} \\
&= \log(-p_t^\pi(r_t - i_t)) && \text{Location shift } i_t \\
&= \log(-p_t^\pi(-\pi_t)) && \text{Fisher's equation} \\
&= \log(p_t^\pi(\pi_t)), && \text{Inversion}
\end{aligned}
$$

   which holds for all $t = 1, ..., T$. $\square$

2. **The PIT histogram for real interest rate density forecasts will be a mirror of the one for inflation density forecasts under the Fisher transformation.**

   We have to prove that $PIT_t^r = 1 - PIT_t^\pi$ for $t = 1, ..., T$. Proof:

$$
\begin{aligned}
PIT_t^r &= P_t^r(r_t) && \text{Definition PIT} \\
&= p(R < r_t) && \text{Definition CDF} \\
&= p(i_t - \Pi < i_t - \pi_t) && \text{Fisher's equation} \\
&= p(-\Pi < -\pi_t) && \text{Substract } i_t \\
&= p(\Pi > \pi_t) && \text{Switch sign} \\
&= 1 - p(\Pi < \pi_t) && \text{Complementary probability} \\
&= 1 - P_t^\pi(\pi_t) && \text{Definition CDF} \\
&= 1 - PIT_t^\pi, && \text{Definition PIT}
\end{aligned}
$$

   which holds for all $t = 1, ..., T$ and where R and Π are random variables for real interest rate and inflation respectively, related by $R = i_t - \Pi$. $\square$

Furthermore, we will prove the following two results to hold under the "Neo-Fisher effect", for which $r_t = i_t + y_t$ such that $p_t^r(\cdot) = i_t + p_t^\pi(\cdot)$ for $t = 1, ..., T$.

3. **The Log Score (LS) is invariant to the Neo-Fisher transformation.**

   We have to prove that $\log(p_t^r(r_t)) = \log(p_t^\pi(y_t))$ for $t = 1, ..., T$. Proof:

$$\begin{aligned}
\log(p_t^r(r_t)) &= \log(i_t + p_t^\pi(r_t)) && \text{Neo-Fisher transformation} \\
&= \log(p_t^\pi(r_t - i_t)) && \text{Location shift } i_t \\
&= \log(p_t^\pi(\pi_t)), && \text{Neo-Fisher's equation}
\end{aligned}$$

   which holds for all $t = 1, ..., T$. $\square$

4. **The PIT histogram for real interest rate density forecasts will be the same as the one for inflation density forecasts under the Neo-Fisher transformation.**

   We have to prove that $PIT_t^r = PIT_t^\pi$ for $t = 1, ..., T$. Proof:

$$\begin{aligned}
PIT_t^r &= P_t^r(r_t) && \text{Definition PIT} \\
&= p(R < r_t) && \text{Definition CDF} \\
&= p(i_t + \Pi < i_t + \pi_t) && \text{Neo-Fisher's equation} \\
&= p(\Pi < \pi_t) && \text{Substract } i_t \\
&= P_t^\pi(\pi_t) && \text{Definition CDF} \\
&= PIT_t^\pi, && \text{Definition PIT}
\end{aligned}$$

   which holds for all $t = 1, ..., T$ and where R and Π are random variables for real interest rate and inflation respectively, related by $R = i_t + \Pi$. $\square$

In this research, we transform to real interest rate density forecasts by applying the Fisher transformation to the "final" combined density forecast for inflation. Moreover, since all four proofs do not depend on combination method, the results hold for both the Linear Pool (LP) and Beta-Transformed Linear Pool (BTLP), such that there is no need to distinct between aggregation technique.

Appendix C Elaboration on Beta-Transformed Linear Pool

The beta distribution for a random variable $X \in [0,1]$ is characterized by the following pdf:

$$b_{\alpha,\beta}(x) = \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \tag{18}$$

where $\mathrm{B}(\alpha,\beta)$ denotes the Beta function given by:

$$\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \tag{19}$$

where $\Gamma(\cdot)$ denotes the Gamma function. Then, the CDF of the beta distribution can be characterized as:

$$B_{\alpha,\beta}(x) = \int_{-\infty}^{x} b_{\alpha,\beta}(s)\, ds. \tag{20}$$

The expectation and variance for this distribution are given by:

$$\mathrm{E}(X) = \frac{\alpha}{\alpha+\beta}, \quad \mathrm{V}(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{21}$$

Below, we present the derivation of Equation 12, moving from the combined CDF of the BTLP to the probability density function. Here, we use that $p(x) = \frac{\partial}{\partial x}P(x)$ (PDF equals derivative of CDF) and the chain rule.

$$\begin{aligned}
\tilde{p}_t(x) &= \frac{\partial}{\partial x}\tilde{P}_t(x) \\
&= \frac{\partial}{\partial x} B_{\alpha,\,\beta}\left(\sum_{k=1}^{K} \mathrm{w}_{k,t} P_{k,t}(x)\right) \\
&= b_{\alpha,\,\beta}\left(\sum_{k=1}^{K} \mathrm{w}_{k,t} P_{k,t}(x)\right) \frac{\partial}{\partial x}\left[\sum_{k=1}^{K} \mathrm{w}_{k,t} P_{k,t}(x)\right] \\
&= b_{\alpha,\,\beta}\left(\sum_{k=1}^{K} \mathrm{w}_{k,t} P_{k,t}(x)\right) \left[\sum_{k=1}^{K} \mathrm{w}_{k,t} \frac{\partial}{\partial x} P_{k,t}(x)\right] \\
&= b_{\alpha,\,\beta}\left(\sum_{k=1}^{K} \mathrm{w}_{k,t} P_{k,t}(x)\right) \left[\sum_{k=1}^{K} \mathrm{w}_{k,t} p_{k,t}(x)\right].
\end{aligned} \tag{22}$$

Finally, we present the algorithm used for the alternating estimation for the BTLP (applied to Simplex regularization) below.

Algorithm 1: Alternating estimation BTLP (Simplex method)

Input: Individual density forecasts $p_{k,t}$ and realizations y_t for $k = 1, ..., K$ and $t = 1, ..., T$.

$\mathbf{w}_1^* = \arg\min_{\mathbf{w}} \sum_{t=1}^{T} -\log\Big(\sum_{k=1}^{K} \mathrm{w}_k p_{k,t}(y_t)\Big)$

s.t. $\mathrm{w}_k \in (0,1)$ for $k = 1, ..., K$ and $\sum_{k=1}^{K} \mathrm{w}_k = 1$;

$\alpha_1^* = \arg\min_{\alpha} \sum_{t=1}^{T} -\log\Big(\big(\sum_{k=1}^{K} \mathrm{w}_{k;1}^* p_{k,t}(y_t)\big) b_{\alpha,1}\big(\sum_{k=1}^{K} \mathrm{w}_{k;1}^* P_{k,t}(y_t)\big)\Big)$;

$\beta_1^* = \arg\min_{\beta} \sum_{t=1}^{T} -\log\Big(\big(\sum_{k=1}^{K} \mathrm{w}_{k;1}^* p_{k,t}(y_t)\big) b_{\alpha_1^*,\beta}\big(\sum_{k=1}^{K} \mathrm{w}_{k;1}^* P_{k,t}(y_t)\big)\Big)$;

for n = 1,..., 5 **do**

$\mathbf{w}_{n+1}^* = \arg\min_{\mathbf{w}} \sum_{t=1}^{T} -\log\Big(\big(\sum_{k=1}^{K} \mathrm{w}_k p_{k,t}(y_t)\big) b_{\alpha_n^*,\beta_n^*}\big(\sum_{k=1}^{K} \mathrm{w}_k P_{k,t}(y_t)\big)\Big)$

s.t. $\mathrm{w}_k \in (0,1)$ for $k = 1, ..., K$ and $\sum_{k=1}^{K} \mathrm{w}_k = 1$;

$\alpha_{n+1}^* = \arg\min_{\alpha} \sum_{t=1}^{T} -\log\Big(\big(\sum_{k=1}^{K} \mathrm{w}_{k;n+1}^* p_{k,t}(y_t)\big) b_{\alpha,\beta_n^*}\big(\sum_{k=1}^{K} \mathrm{w}_{k;n+1}^* P_{k,t}(y_t)\big)\Big)$;

$\beta_{n+1}^* = \arg\min_{\beta} \sum_{t=1}^{T} -\log\Big(\big(\sum_{k=1}^{K} \mathrm{w}_{k;n+1}^* p_{k,t}(y_t)\big) b_{\alpha_{n+1}^*,\beta}\big(\sum_{k=1}^{K} \mathrm{w}_{k;n+1}^* P_{k,t}(y_t)\big)\Big)$;

end

Output: Estimates of optimal weights, α and β for the BTLP (Simplex method).

Appendix D Additional Figures

Appendix D1 Combined densities regularized mixtures for Eurozone inflation

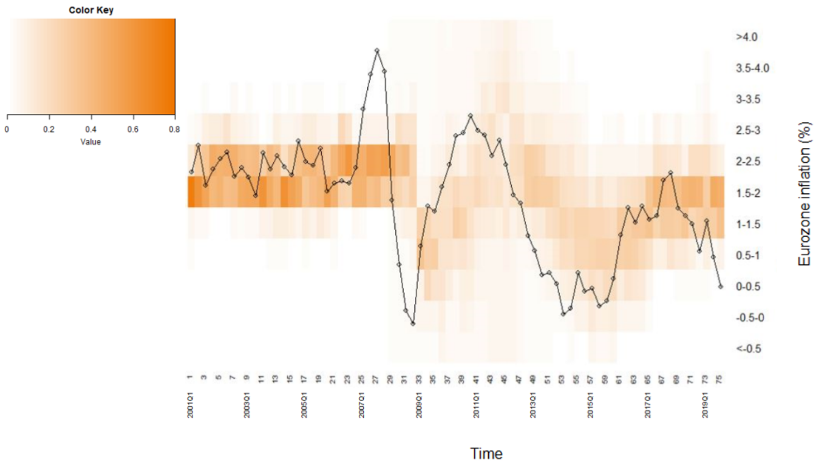


Figure 22: Simplex combined density forecasts for (superimposed) Eurozone inflation 2001Q1 - 2019Q3, displayed as heatmap.

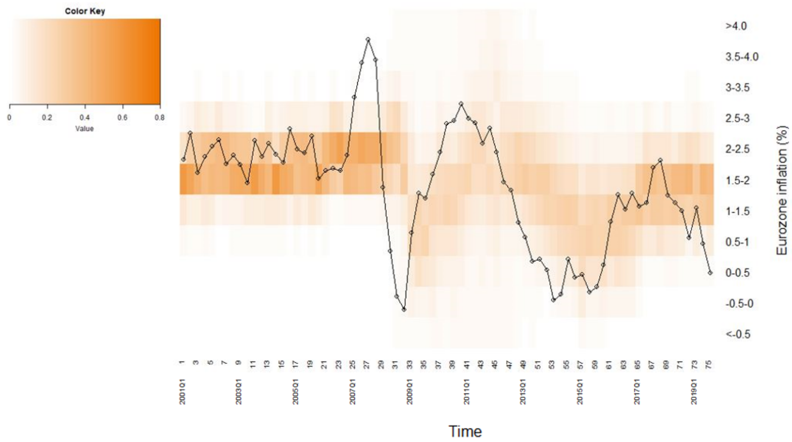


Figure 23: Simplex+Entropy combined density forecasts for (superimposed) Eurozone inflation 2001Q1 - 2019Q3, displayed as heatmap.

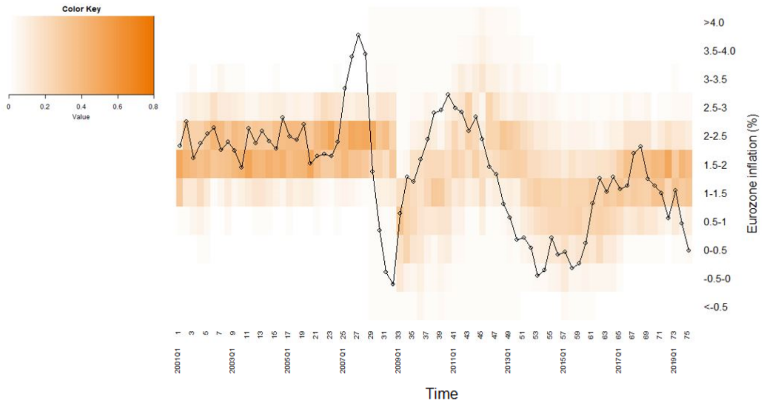


Figure 24: Best 4-Average combined density forecasts for (superimposed) Eurozone inflation 2001Q1 - 2019Q3, displayed as heatmap.

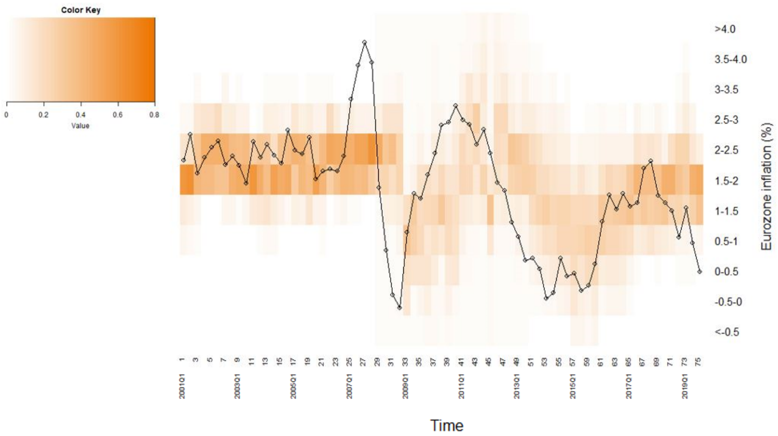


Figure 25: Best ≤4-Average combined density forecasts for (superimposed) Eurozone inflation 2001Q1 - 2019Q3, displayed as heatmap.

Appendix D2 Heatmaps differences Simple Average and regularized density forecasts for Eurozone inflation

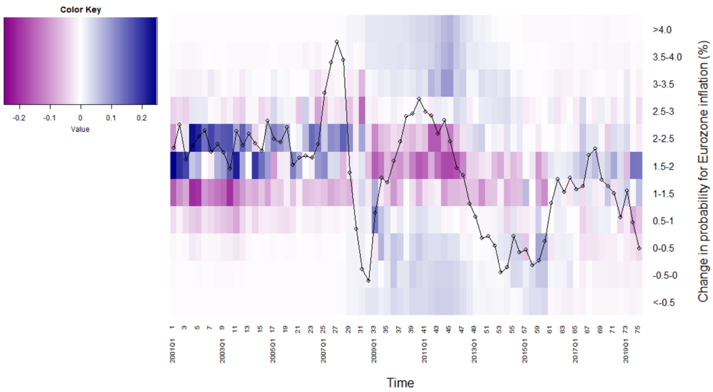


Figure 26: Heat map of differences between Simplex and Simple Average combined density forecasts (Simplex minus Simple Average). Realized inflation is superimposed.

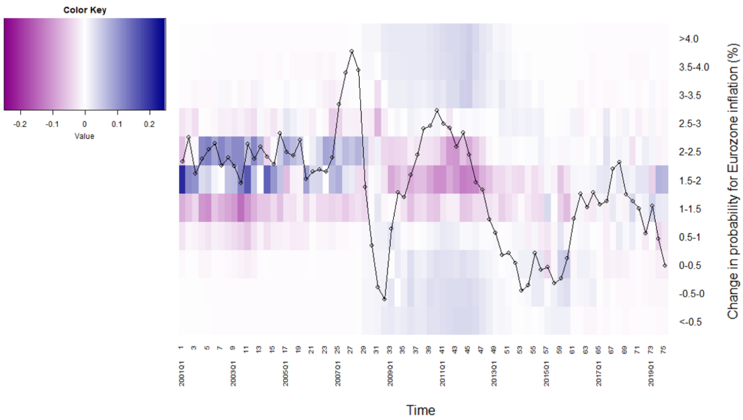


Figure 27: Heat map of differences between Simplex+Entropy and Simple Average combined density forecasts (Simplex+Entropy minus Simple Average). Realized inflation is superimposed.

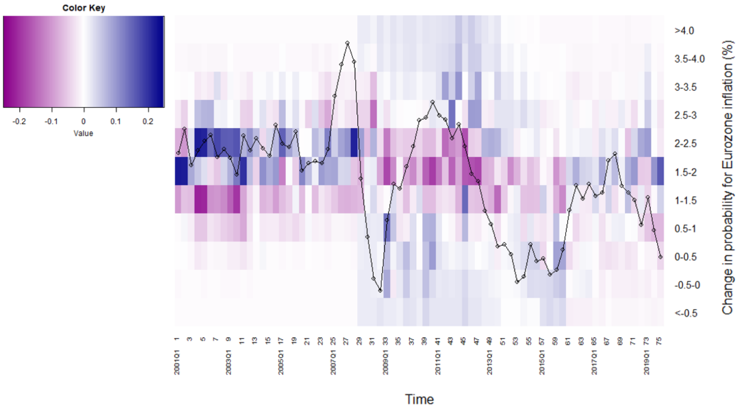


Figure 28: Heat map of differences between Best 4-Average and Simple Average combined density forecasts (Best 4-Average minus Simple Average). Realized inflation is superimposed.

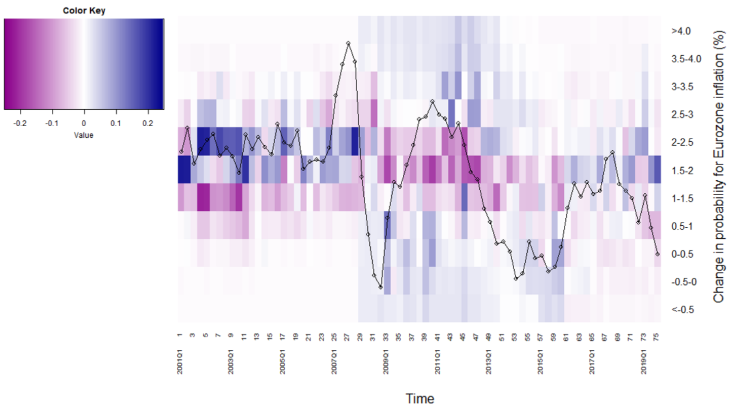
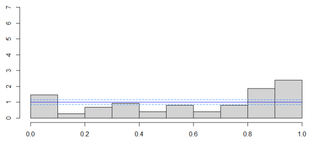
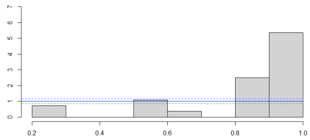


Figure 29: Heat map of differences between Best ≤4-Average and Simple Average combined density forecasts (Best ≤4-Average minus Simple Average). Realized inflation is superimposed.

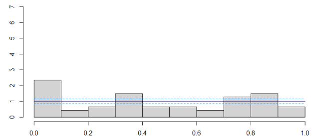
Appendix D3 PIT histograms regularized density forecasts for Eurozone inflation



(a) Full sample

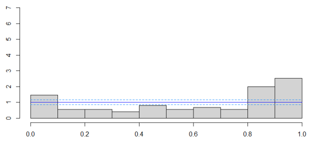


(b) First subsample, 2001Q1 - 2007Q4

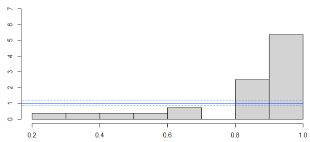


(c) Second subsample, 2008Q1 - 2019Q3

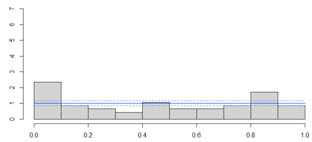
Figure 30: Histograms of the PIT series for the Simplex method, Eurozone inflation 2001Q1-2019Q3. 95% confidence intervals are dotted in light blue.



(a) Full sample

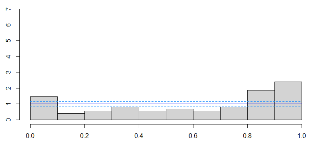


(b) First subsample, 2001Q1 - 2007Q4

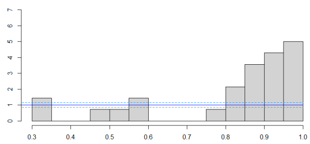


(c) Second subsample, 2008Q1 - 2019Q3

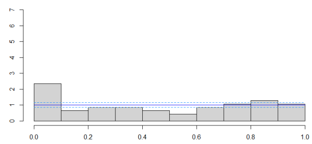
Figure 31: Histograms of the PIT series for the Simplex+Entropy method, Eurozone inflation 2001Q1-2019Q3. 95% confidence intervals are dotted in light blue.



(a) Full sample

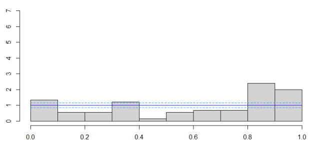


(b) First subsample, 2001Q1 - 2007Q4

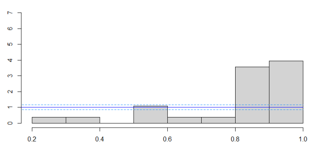


(c) Second subsample, 2008Q1 - 2019Q3

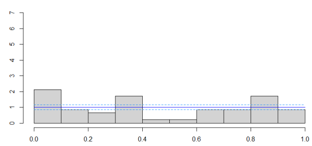
Figure 32: Histograms of the PIT series for the Best 4-Average method, Eurozone inflation 2001Q1-2019Q3. 95% confidence intervals are dotted in light blue.



(a) Full sample



(b) First subsample, 2001Q1 - 2007Q4



(c) Second subsample, 2008Q1 - 2019Q3

Figure 33: Histograms of the PIT series for the Best $\leq$4-Average method, Eurozone inflation 2001Q1-2019Q3. 95% confidence intervals are dotted in light blue.

Appendix D4 Combined densities regularized mixtures for Eurozone real interest rate

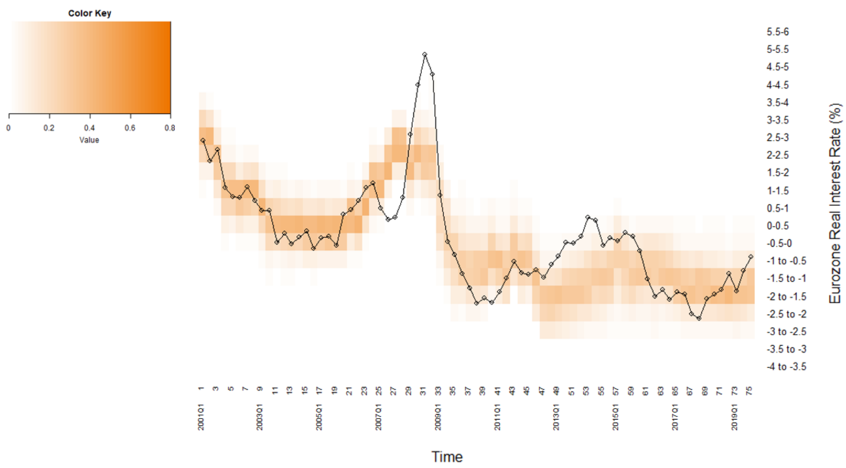


Figure 34: Simple Average combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1 - 2019Q3 (75 quarters), displayed as heatmap.

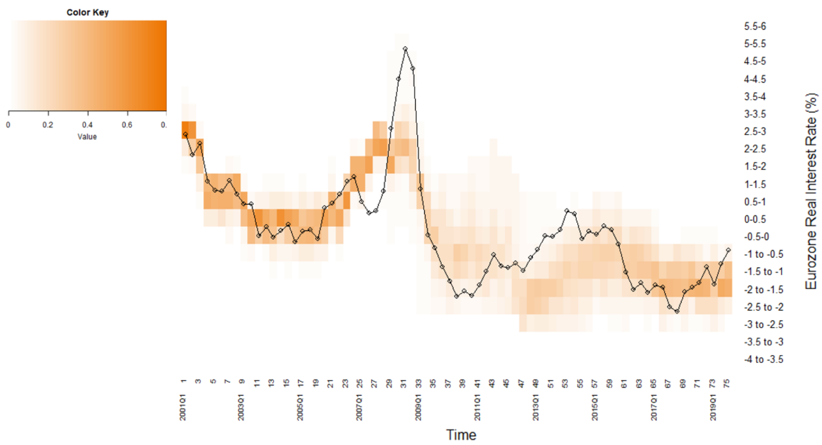


Figure 35: Simplex combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1 - 2019Q3 (75 quarters), displayed as heatmap.

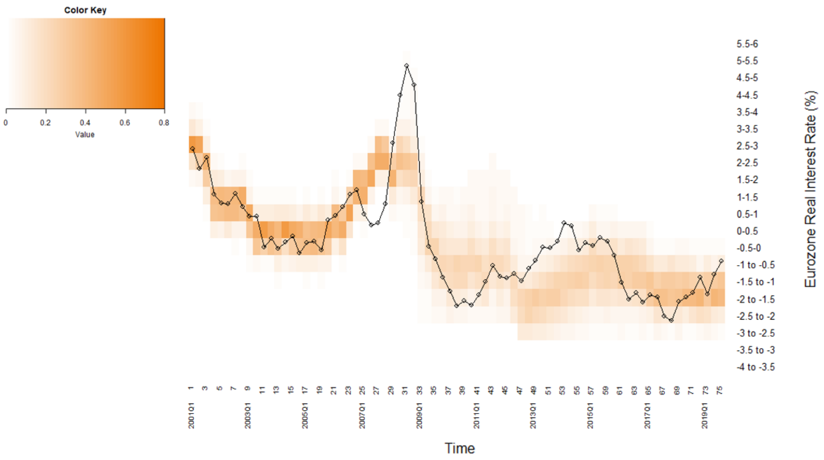


Figure 36: Simplex+Entropy combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1 - 2019Q3 (75 quarters), displayed as heatmap.

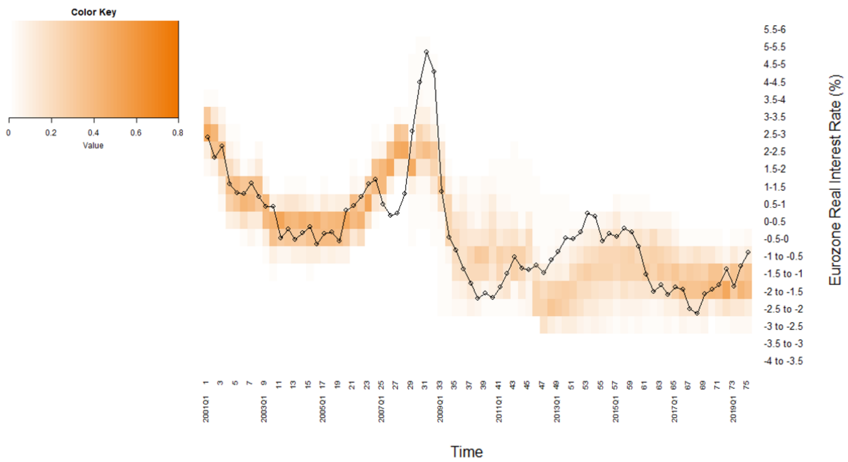


Figure 37: Best 4-Average combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1 - 2019Q3 (75 quarters), displayed as heatmap.

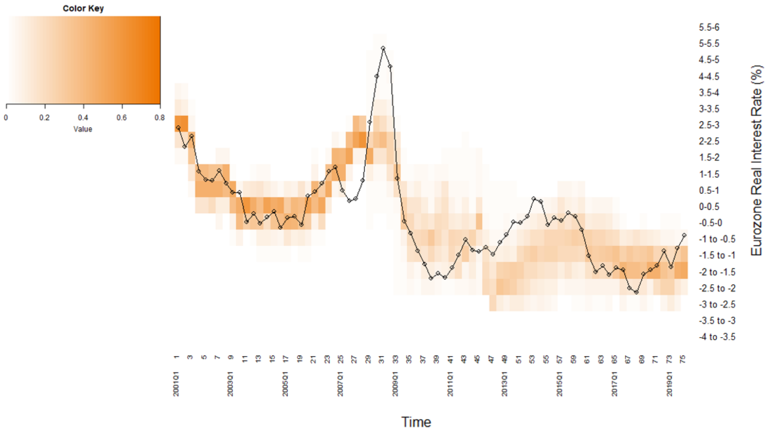


Figure 38: Best ≤4-Average combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1 - 2019Q3 (75 quarters), displayed as heatmap.

Appendix D5 Heatmaps differences Simple Average and regularized density forecasts, Eurozone real interest rate

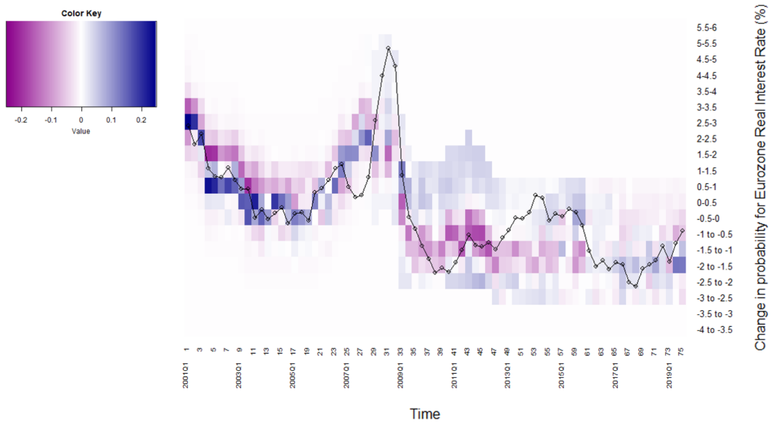


Figure 39: Heat map of differences between Simplex and Simple Average combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1-2019Q3 (Simplex minus Simple Average).

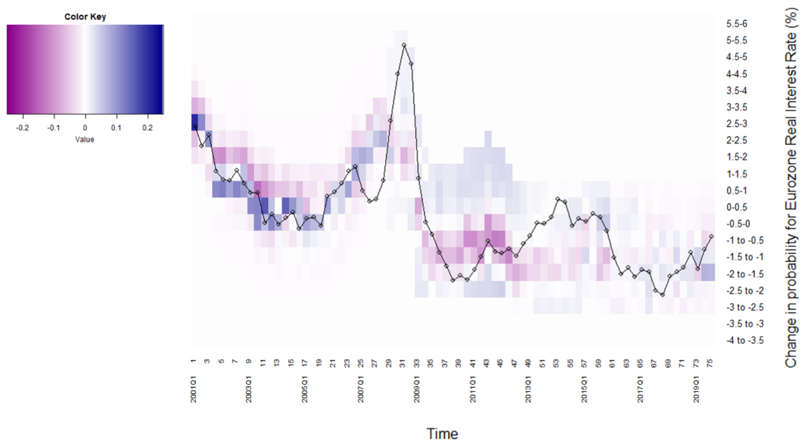


Figure 40: Heat map of differences between Simplex+Entropy and Simple Average combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1-2019Q3 (Simplex+Entropy minus Simple Average).

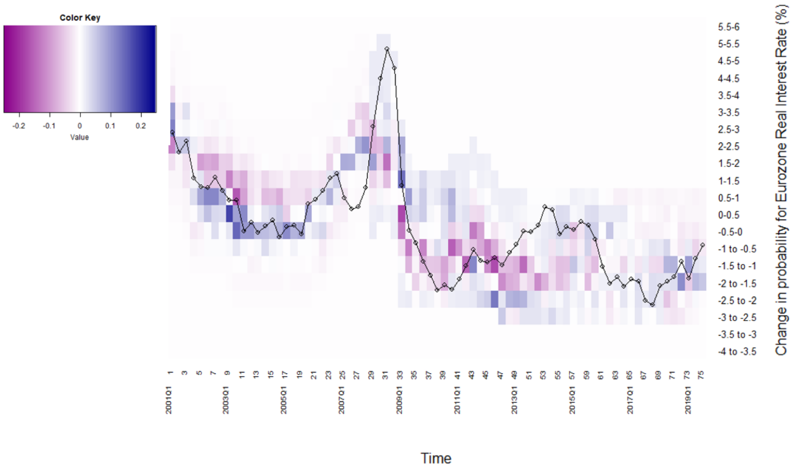


Figure 41: Heat map of differences between Simplex+Ridge and Simple Average combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1-2019Q3 (Best 4-Average minus Simple Average).

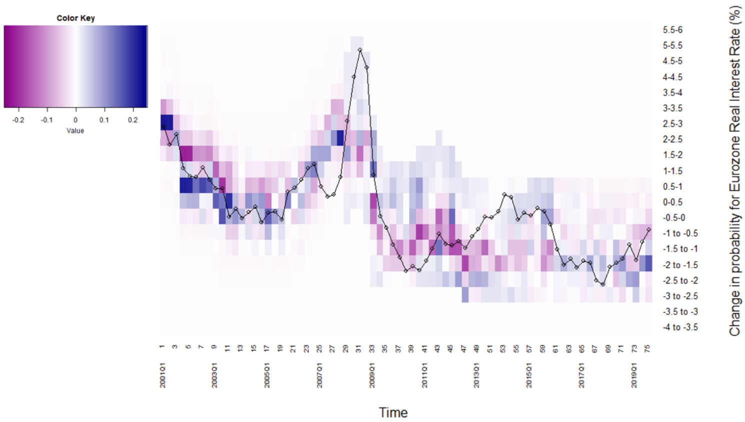
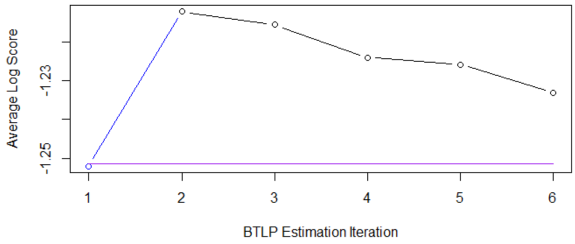
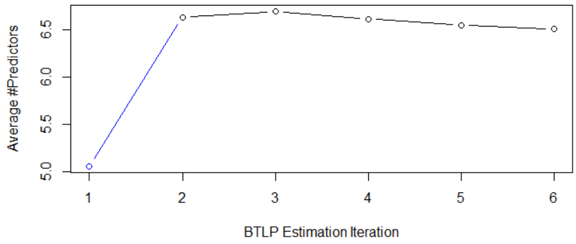


Figure 42: Heat map of differences between Simplex+Ridge and Simple Average combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1-2019Q3 (Best ≤4-Average minus Simple Average).

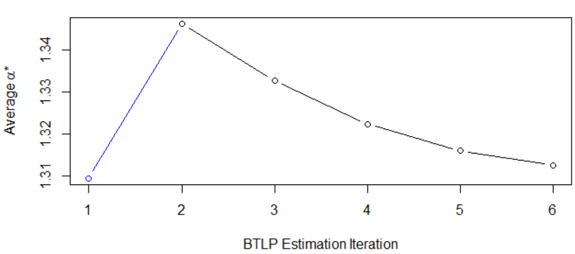
Appendix D6 BTLP alternating estimation graphs simulation



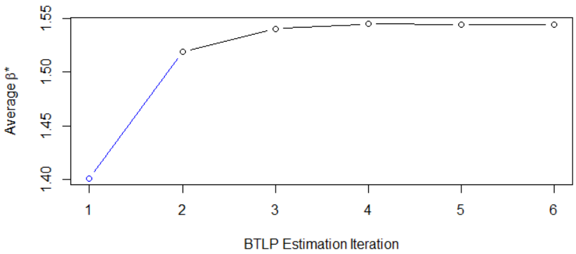
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors

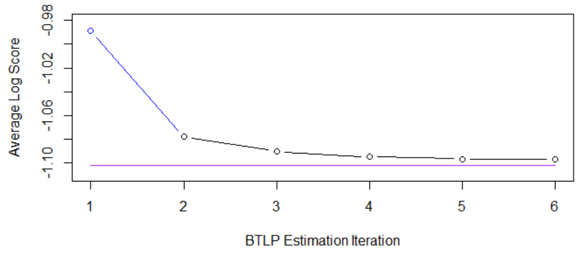


(c) Average α^*

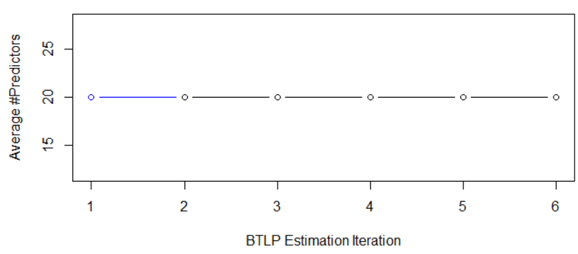


(d) Average β^*

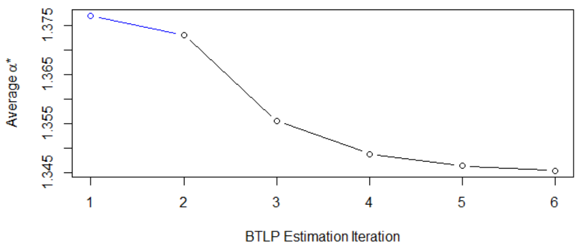
Figure 43: Simulation results (Log score, #Predictors, α^*, β^*) for the BTLP with respect to estimation iteration, for Simplex, DGP 1. First "One-time" estimation in blue, subsequent alternating estimations in black.



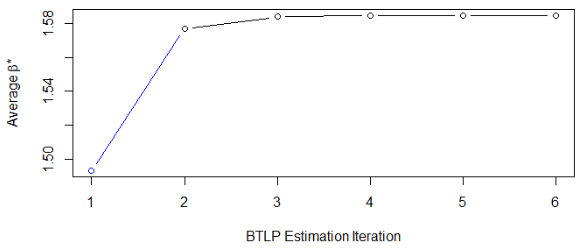
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors

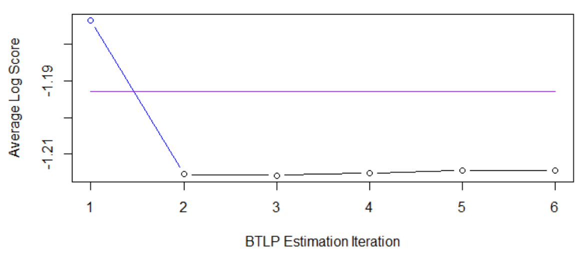


(c) Average α^*

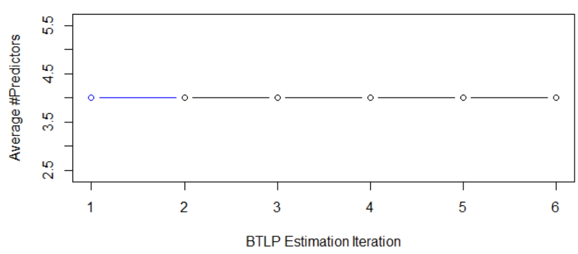


(d) Average β^*

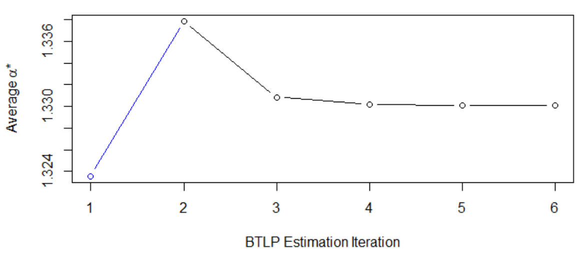
Figure 44: Simulation results (Log score, #Predictors, α^*, β^*) for the BTLP with respect to estimation iteration, for Simplex+Entropy, DGP 1. First "One-time" estimation in blue, subsequent alternating estimations in black.



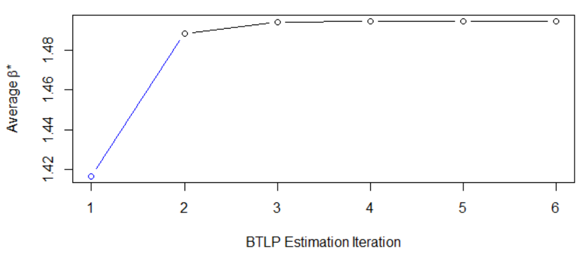
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors

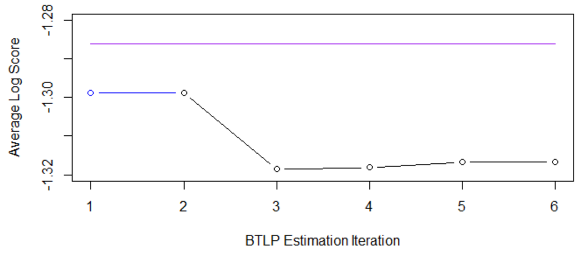


(c) Average α^*

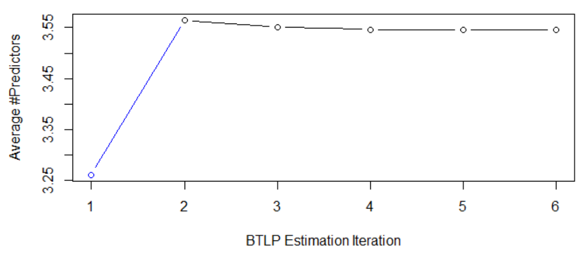


(d) Average β^*

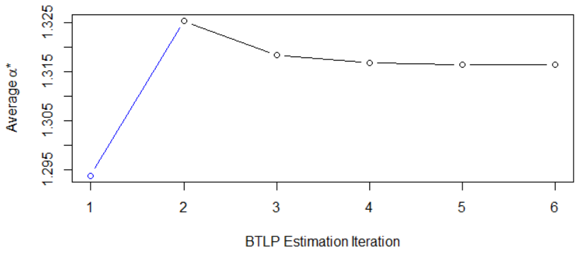
Figure 45: Simulation results (Log score, #Predictors, α^*, β^*) for the BTLP with respect to estimation iteration, for Best 4-Average, DGP 1. First "One-time" estimation in blue, subsequent alternating estimations in black.



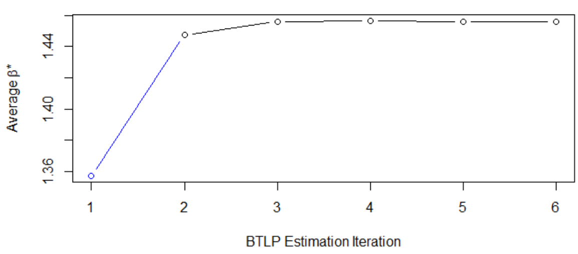
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors

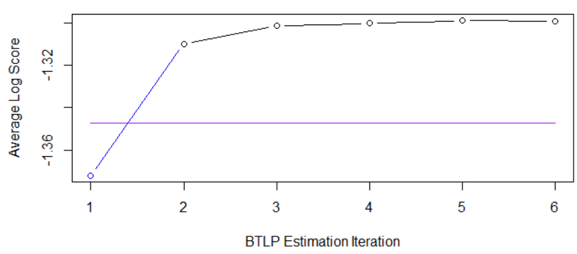


(c) Average α^*

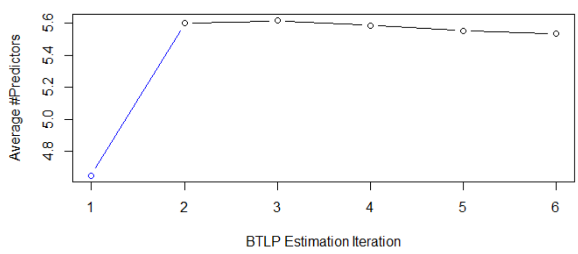


(d) Average β^*

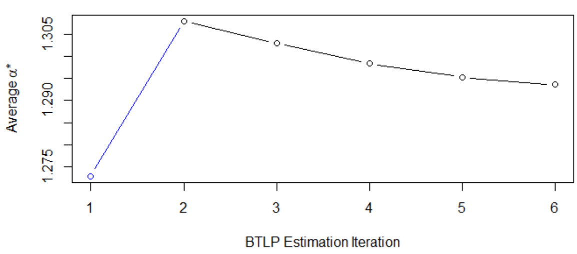
Figure 46: Simulation results (Log score, #Predictors, α^*, β^*) for the BTLP with respect to estimation iteration, for Best ≤4-Average, DGP 1. First "One-time" estimation in blue, subsequent alternating estimations in black.



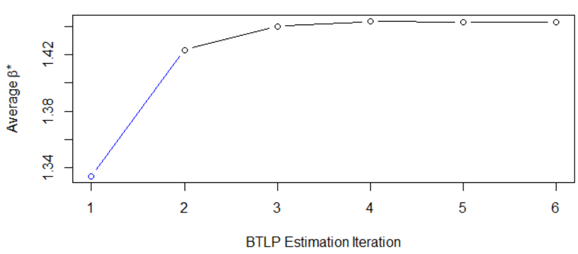
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors

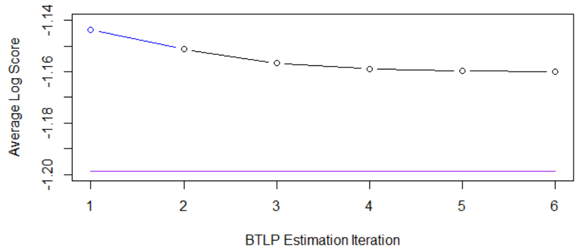


(c) Average α^*

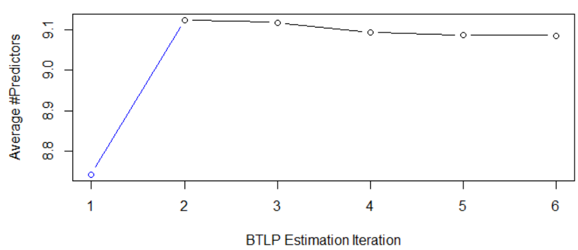


(d) Average β^*

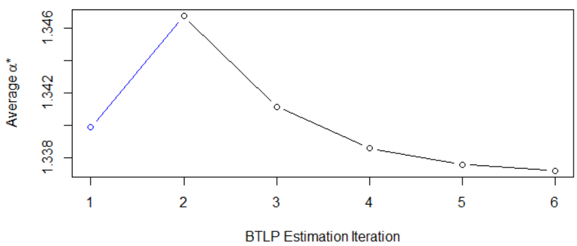
Figure 47: Simulation results (Log score, #Predictors, α^*, β^*) for the BTLP with respect to estimation iteration, for Simplex, DGP 2. First "One-time" estimation in blue, subsequent alternating estimations in black.



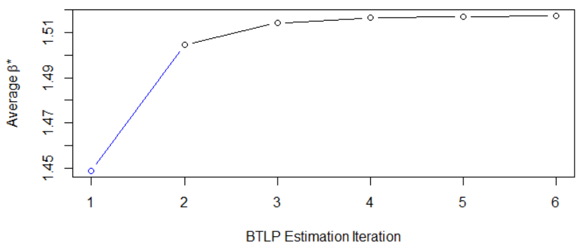
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors

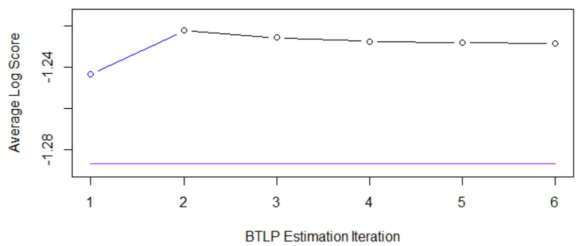


(c) Average α^*

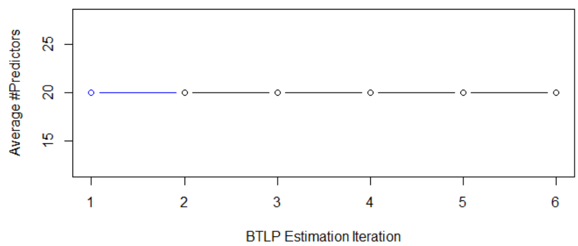


(d) Average β^*

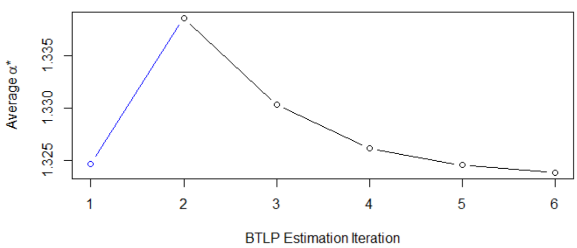
Figure 48: Simulation results (Log score, #Predictors, α^*, β^*) for the BTLP with respect to estimation iteration, for Simplex+Ridge, DGP 2. First "One-time" estimation in blue, subsequent alternating estimations in black.



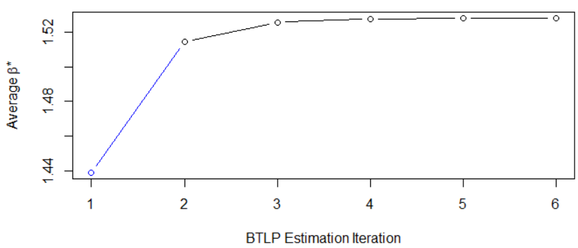
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors

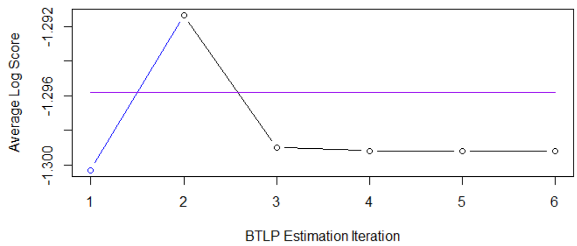


(c) Average α^*

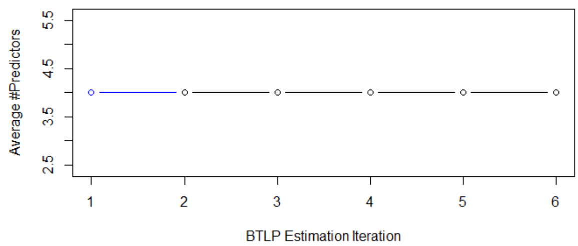


(d) Average β^*

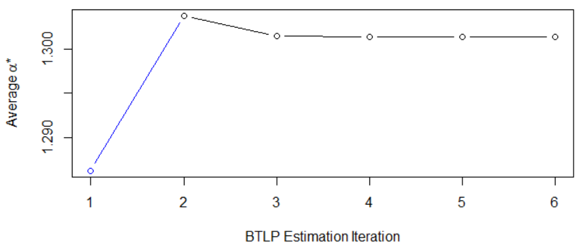
Figure 49: Simulation results (Log score, #Predictors, α^*, β^*) for the BTLP with respect to estimation iteration, for Simplex+Entropy, DGP 2. First "One-time" estimation in blue, subsequent alternating estimations in black.



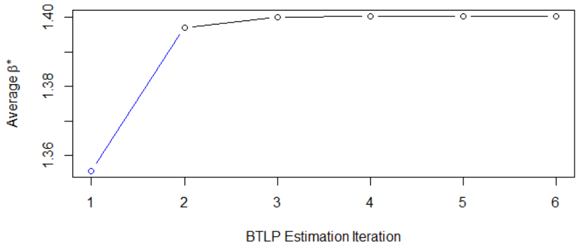
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors

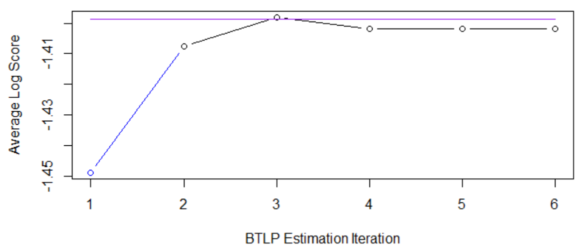


(c) Average α^*

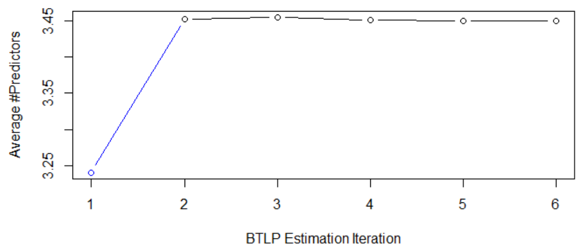


(d) Average β^*

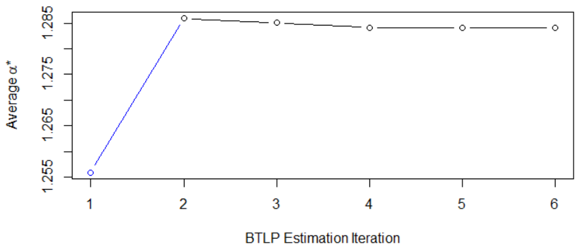
Figure 50: Simulation results (Log score, #Predictors, α^*, β^*) for the BTLP with respect to estimation iteration, for Best 4-Average, DGP 2. First "One-time" estimation in blue, subsequent alternating estimations in black.



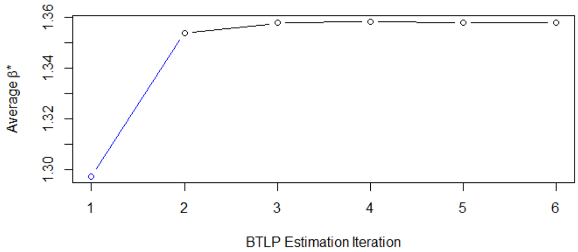
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors



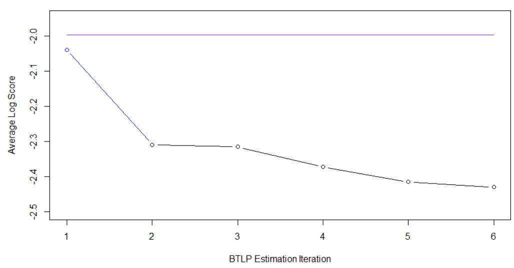
(c) Average α^*



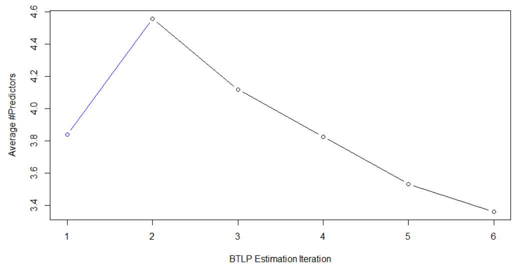
(d) Average β^*

Figure 51: Simulation results (Log score, #Predictors, α^*, β^*) for the BTLP with respect to estimation iteration, for Best $\leq$4-Average, DGP 2. First "One-time" estimation in blue, subsequent alternating estimations in black.

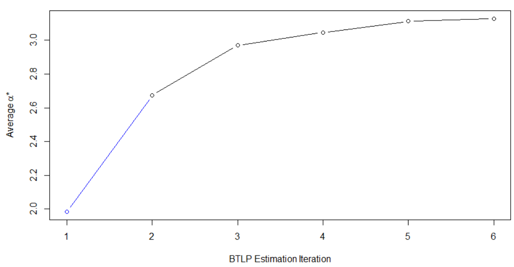
Appendix D7 BTLP alternating estimation graphs Eurozone inflation



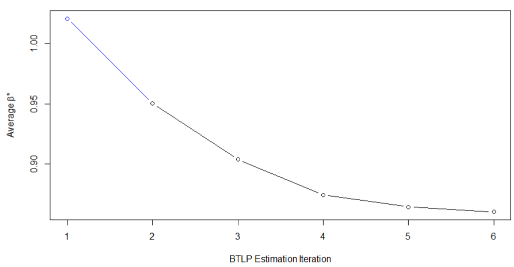
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors

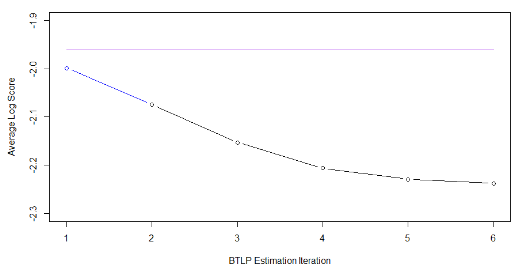


(c) Average α^*

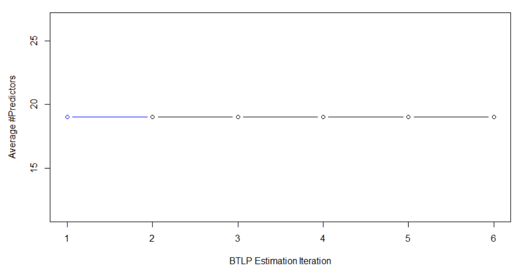


(d) Average β^*

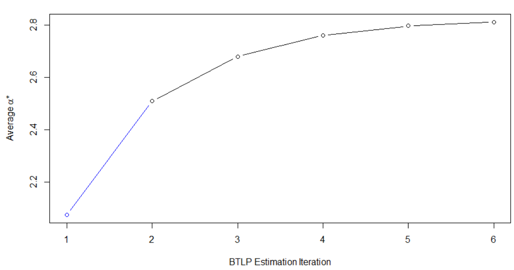
Figure 52: Empirical results (Log score, #Predictors, α^*, β^*) for the BTLP with respect to estimation iteration, for Simplex, Eurozone inflation. First "One-time" estimation in blue, subsequent alternating estimations in black.



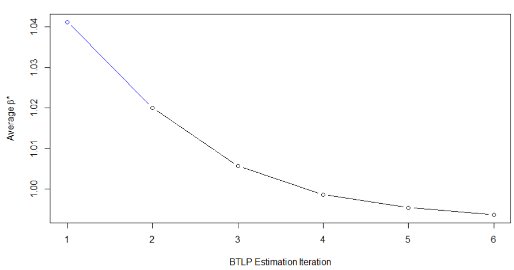
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors

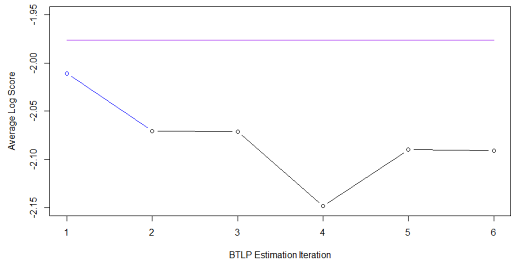


(c) Average α^*

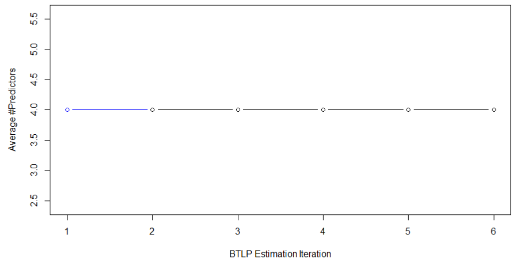


(d) Average β^*

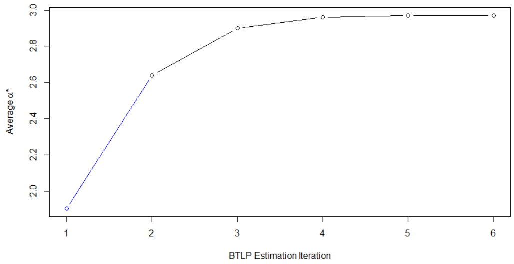
Figure 53: Empirical results (Log score, #Predictors, α^*, β^*) for the BTLP with respect to estimation iteration, for Simplex+Entropy, Eurozone inflation. First "One-time" estimation in blue, subsequent alternating estimations in black.



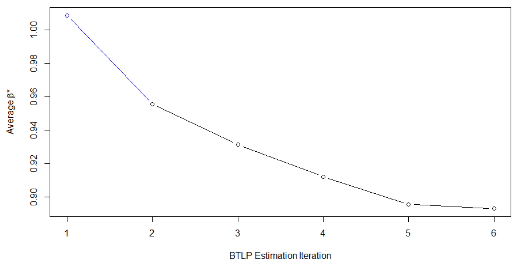
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors

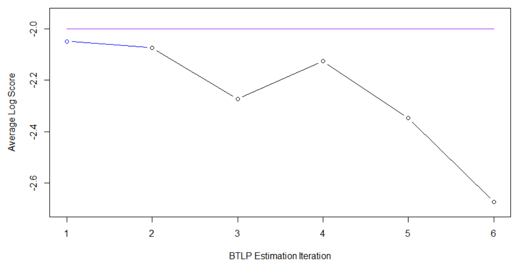


(c) Average α^*

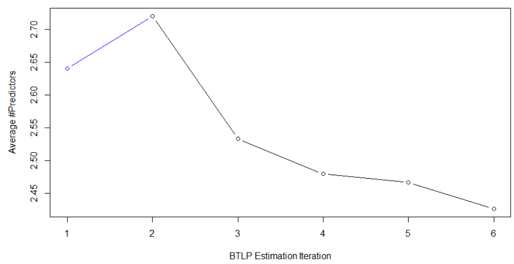


(d) Average β^*

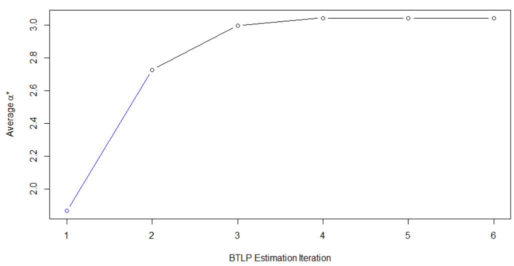
Figure 54: Empirical results (Log score, #Predictors, α^*, β^*) for the BTLP with respect to estimation iteration, for Best 4-Average, Eurozone inflation. First "One-time" estimation in blue, subsequent alternating estimations in black.



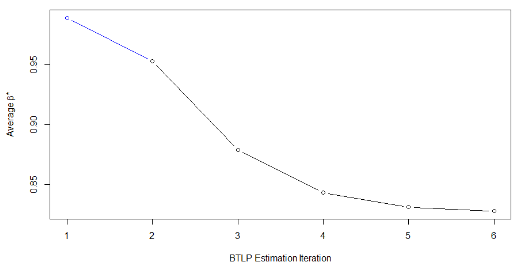
(a) Average Log Score (Linear Pool in purple)



(b) Average #Predictors



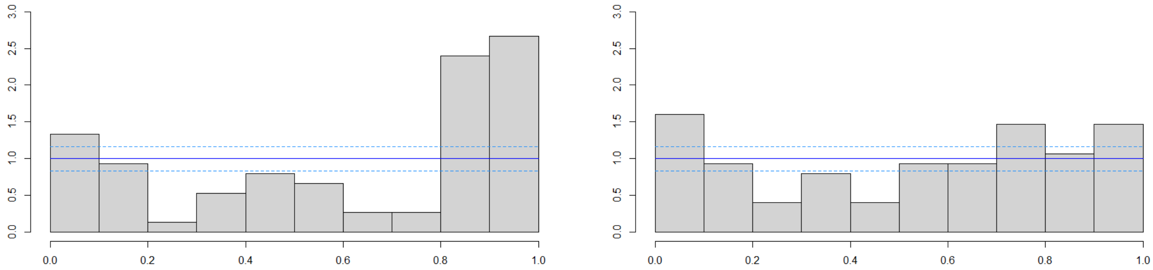
(c) Average α^*



(d) Average β^*

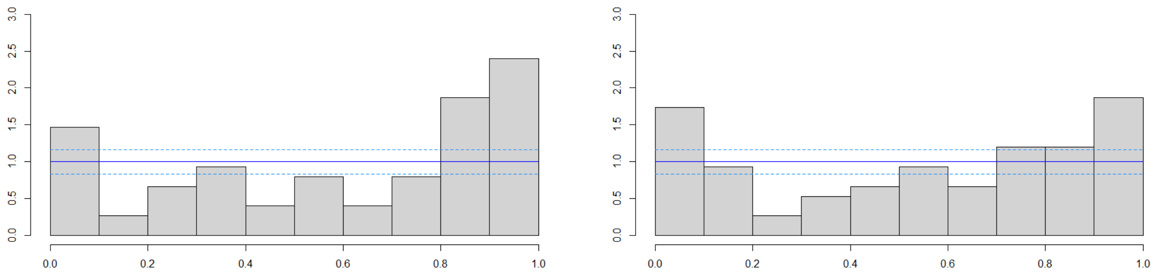
Figure 55: Empirical results (Log score, #Predictors, α^*, β^*) for the BTLP with respect to estimation iteration, for Best $\leq$4-Average, Eurozone inflation. First "One-time" estimation in blue, subsequent alternating estimations in black.

Appendix D8 BTLP PIT histograms regularized density forecasts for Eurozone inflation



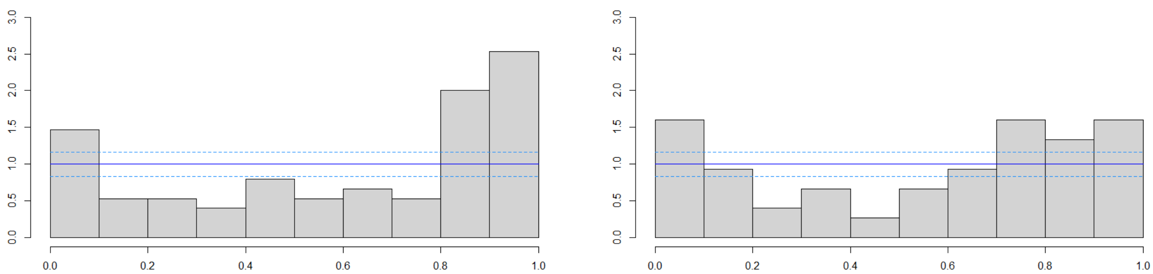
(a) Linear Pool **(b)** Beta-Transformed Linear Pool

Figure 56: PIT histograms Simple Average for Eurozone inflation, using Linear Pool and Beta-Transformed Linear Pool ("One-Time" Estimation). 95% confidence intervals are dotted in light blue.



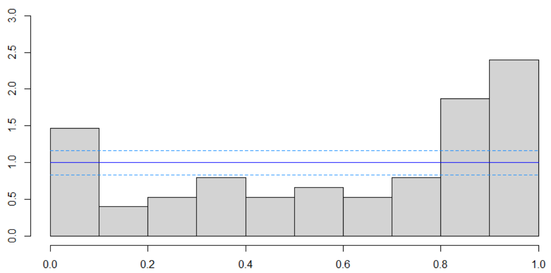
(a) Linear Pool **(b)** Beta-Transformed Linear Pool

Figure 57: PIT histograms Simplex for Eurozone inflation, using Linear Pool and Beta-Transformed Linear Pool ("One-Time" Estimation). 95% confidence intervals are dotted in light blue.

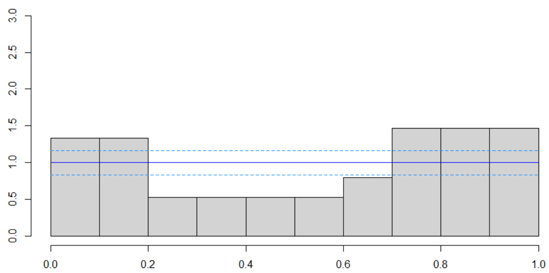


(a) Linear Pool **(b)** Beta-Transformed Linear Pool

Figure 58: PIT histograms Simplex+Entropy for Eurozone inflation, using Linear Pool and Beta-Transformed Linear Pool ("One-Time" Estimation). 95% confidence intervals are dotted in light blue.

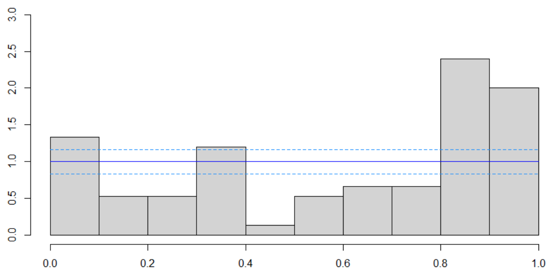


(a) Linear Pool

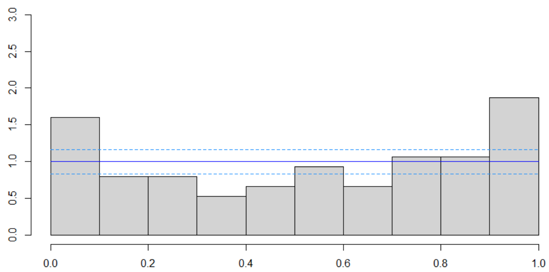


(b) Beta-Transformed Linear Pool

Figure 59: PIT histograms Best 4-Average for Eurozone inflation, using Linear Pool and Beta-Transformed Linear Pool ("One-Time" Estimation). 95% confidence intervals are dotted in light blue.



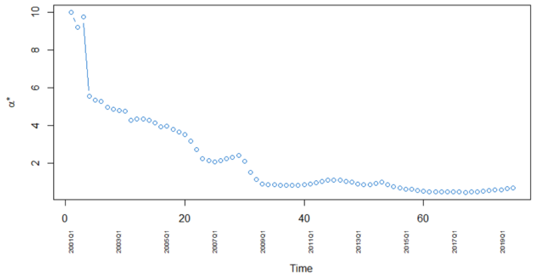
(a) Linear Pool



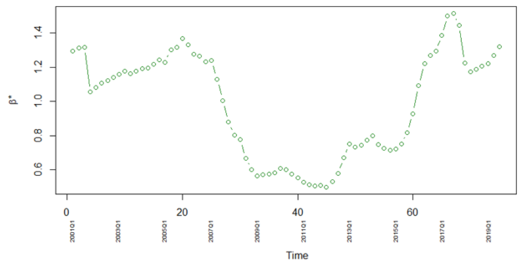
(b) Beta-Transformed Linear Pool

Figure 60: PIT histograms Best $\leq$4-Average for Eurozone inflation, using Linear Pool and Beta-Transformed Linear Pool ("One-Time" Estimation). 95% confidence intervals are dotted in light blue.

Appendix D9 BTLP estimates of α and β over time for regularized density forecasts for Eurozone inflation

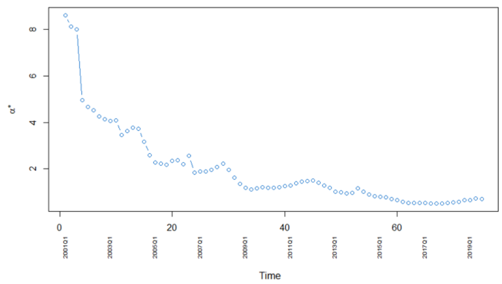


(a) α^*

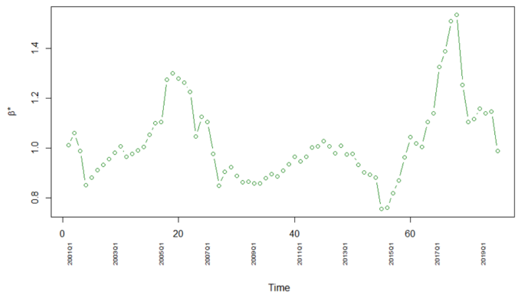


(b) β^*

Figure 61: "One-time" estimates of α^* and β^* for the BTLP over time, for Simple Average, Eurozone inflation.

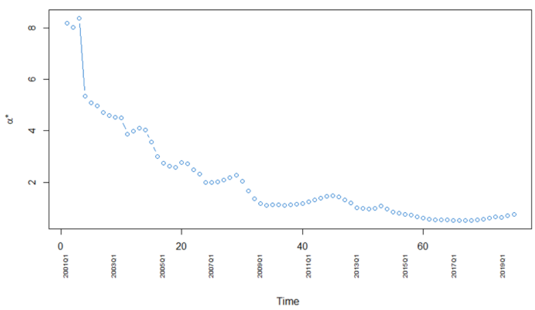


(a) α^*

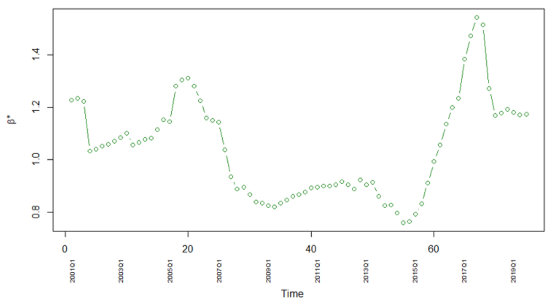


(b) β^*

Figure 62: "One-time" estimates of α^* and β^* for the BTLP over time, for Simplex, Eurozone inflation.

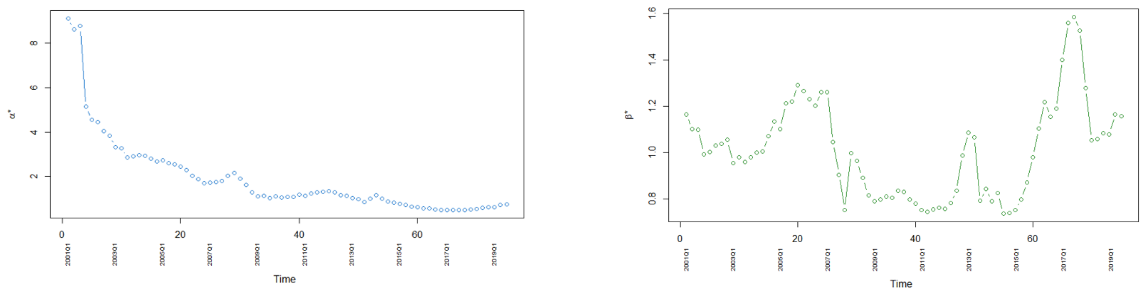


(a) α^*



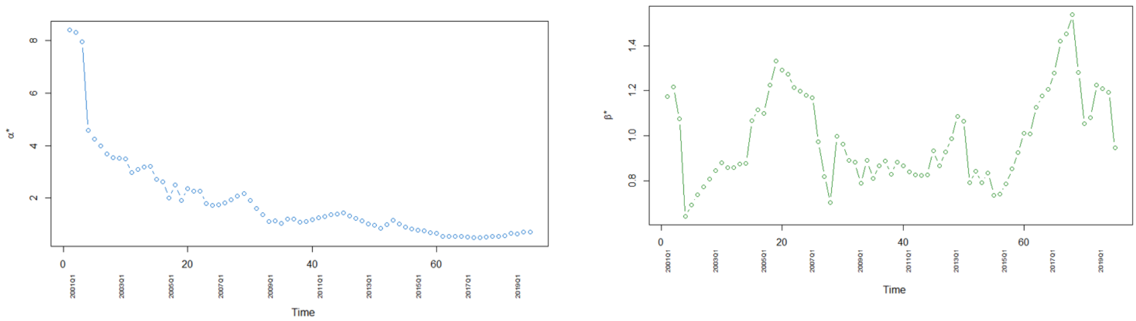
(b) β^*

Figure 63: "One-time" estimates of α^* and β^* for the BTLP over time, for Simplex+Entropy, Eurozone inflation.



(a) α^* **(b)** β^*

Figure 64: "One-time" estimates of α^* and β^* for the BTLP over time, for Best 4-Average, Eurozone inflation.



(a) α^* **(b)** β^*

Figure 65: "One-time" estimates of α^* and β^* for the BTLP over time, for Best $\leq$4-Average, Eurozone inflation.

Appendix D10 BTLP combined densities regularized mixtures for Eurozone inflation

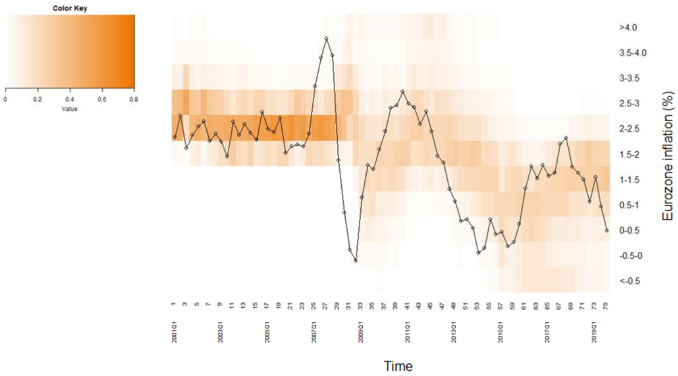


Figure 66: Simple Average combined density forecasts for (superimposed) Eurozone inflation 2001Q1 - 2019Q3 for the BTLP ("One-Time" Estimation), displayed as heatmap.

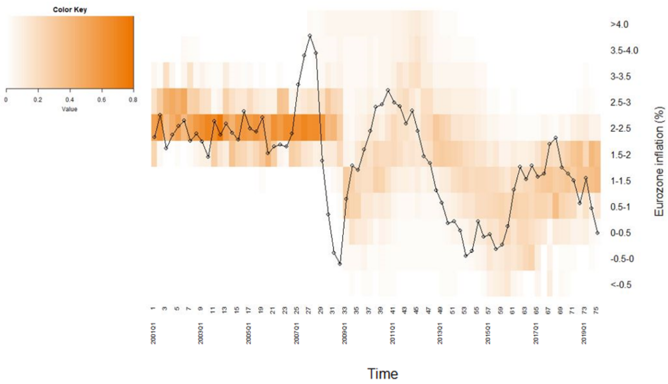


Figure 67: Simplex combined density forecasts for (superimposed) Eurozone inflation 2001Q1 - 2019Q3 for the BTLP ("One-Time" Estimation), displayed as heatmap.

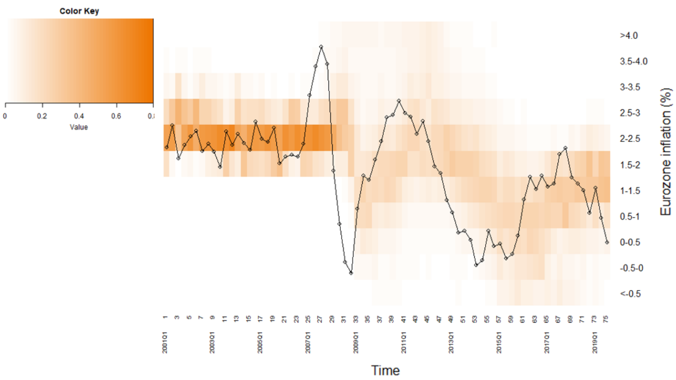


Figure 68: Simplex+Entropy combined density forecasts for (superimposed) Eurozone inflation 2001Q1 - 2019Q3 for the BTLP ("One-Time" Estimation), displayed as heatmap.

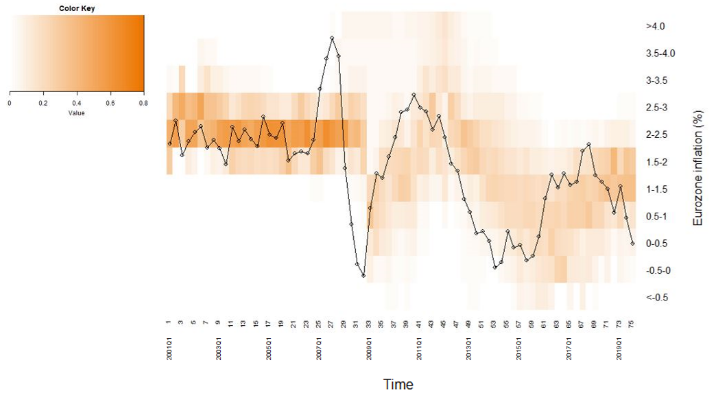


Figure 69: Best 4-Average combined density forecasts for (superimposed) Eurozone inflation 2001Q1 - 2019Q3 for the BTLP ("One-Time" Estimation), displayed as heatmap.

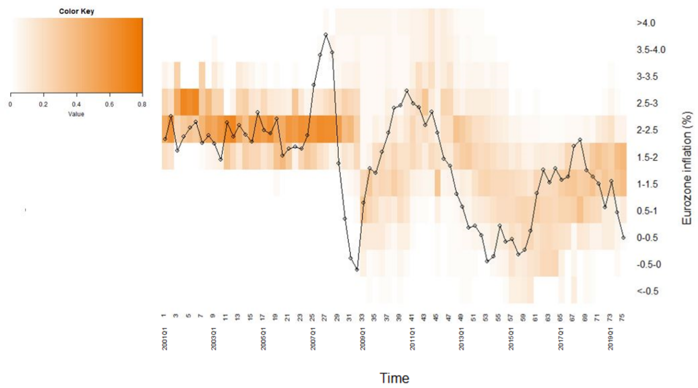


Figure 70: Best ≤4-Average combined density forecasts for (superimposed) Eurozone inflation 2001Q1 - 2019Q3 for the BTLP ("One-Time" Estimation), displayed as heatmap.

Appendix D11 Heatmaps differences LP and BTLP of regularized density forecasts for Eurozone inflation

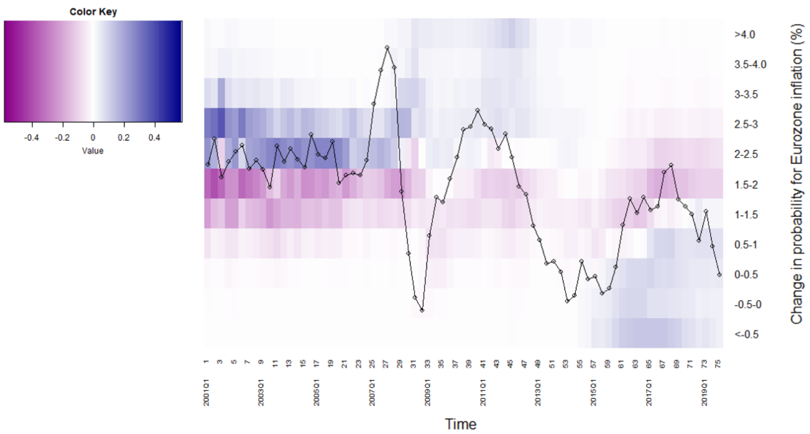


Figure 71: Heat map of differences between BTLP and LP for Simple Average combined density forecasts for (superimposed) Eurozone inflation 2001Q1-2019Q3 (BTLP minus LP).

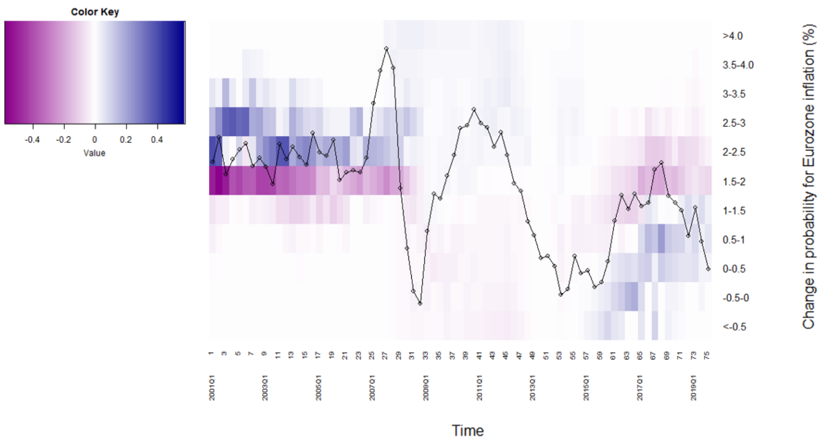


Figure 72: Heat map of differences between BTLP and LP for Simplex combined density forecasts for (superimposed) Eurozone inflation 2001Q1-2019Q3 (BTLP minus LP).

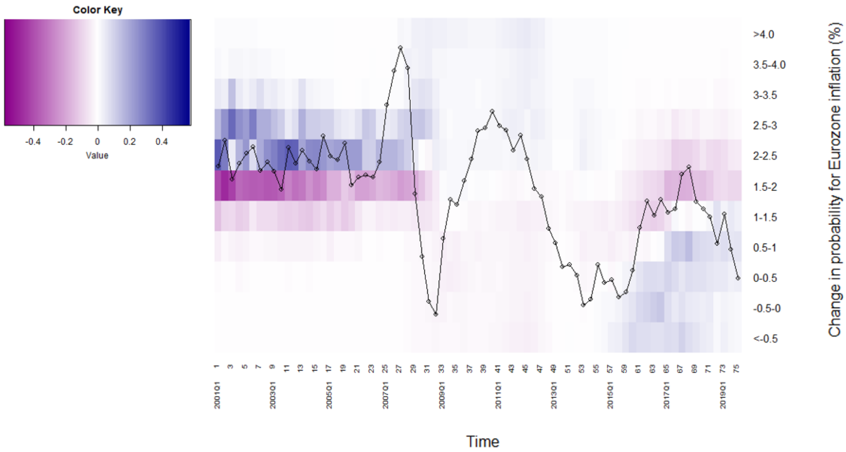


Figure 73: Heat map of differences between BTLP and LP for Simplex+Entropy combined density forecasts for (superimposed) Eurozone inflation 2001Q1-2019Q3 (BTLP minus LP).

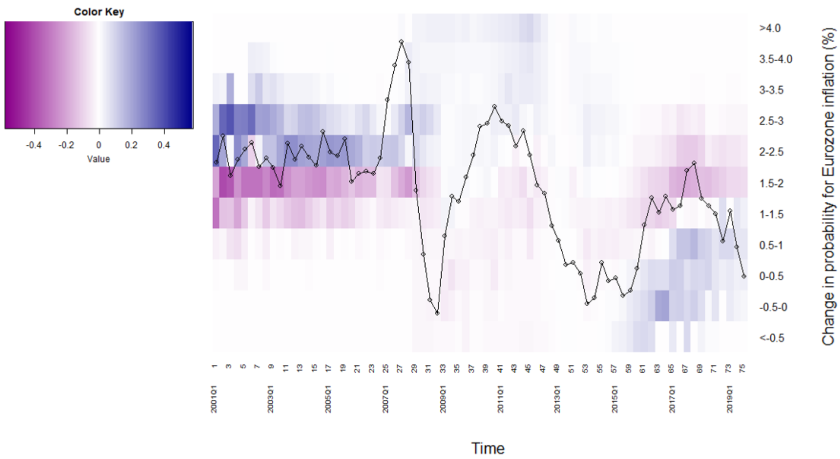


Figure 74: Heat map of differences between BTLP and LP for Best 4-Average combined density forecasts for (superimposed) Eurozone inflation 2001Q1-2019Q3 (BTLP minus LP).

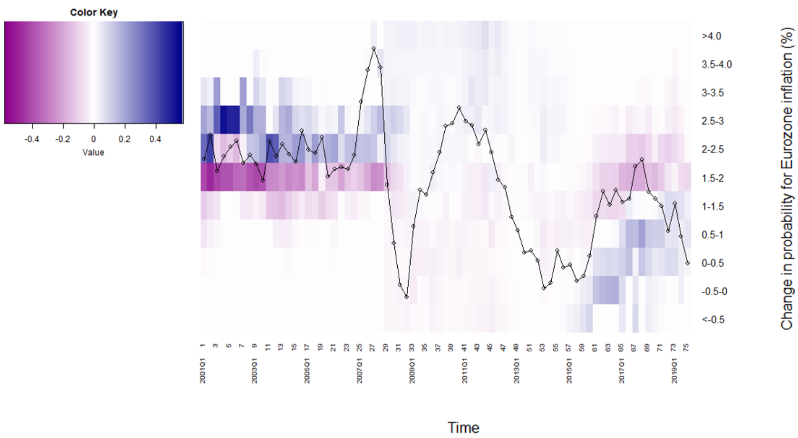


Figure 75: Heat map of differences between BTLP and LP for Best ≤4-Average combined density forecasts for (superimposed) Eurozone inflation 2001Q1-2019Q3 (BTLP minus LP).

Appendix D12 BTLP combined densities regularized mixtures for Eurozone real interest rate

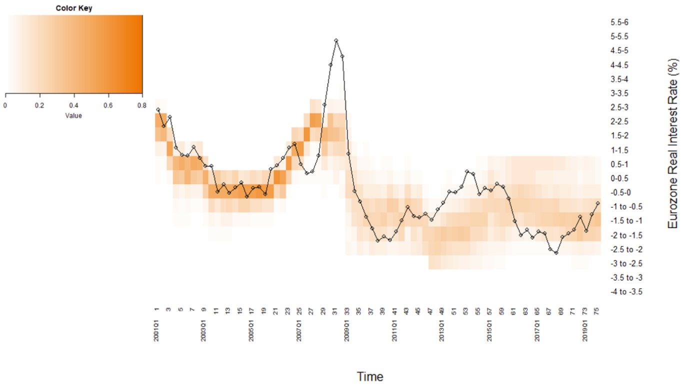


Figure 76: Simple Average combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1 - 2019Q3 for the BTLP ("One-Time" Estimation), displayed as heatmap.

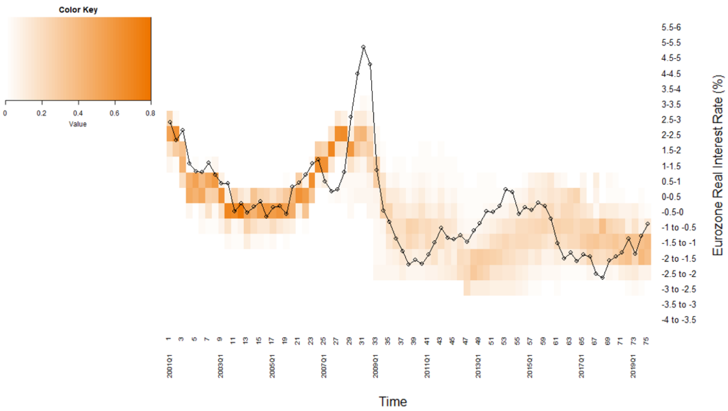


Figure 77: Simplex combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1 - 2019Q3 for the BTLP ("One-Time" Estimation), displayed as heatmap.

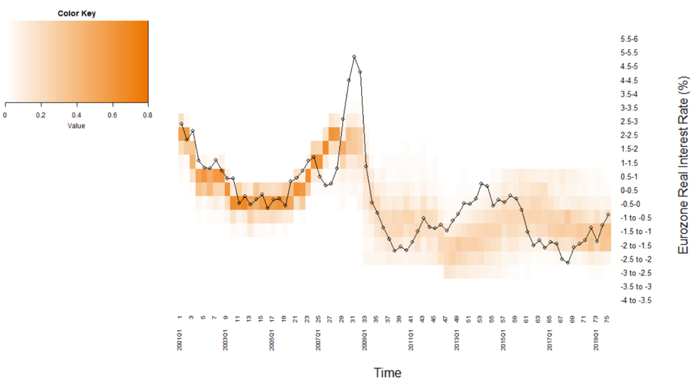


Figure 78: Simplex+Entropy combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1 - 2019Q3 for the BTLP ("One-Time" Estimation), displayed as heatmap.

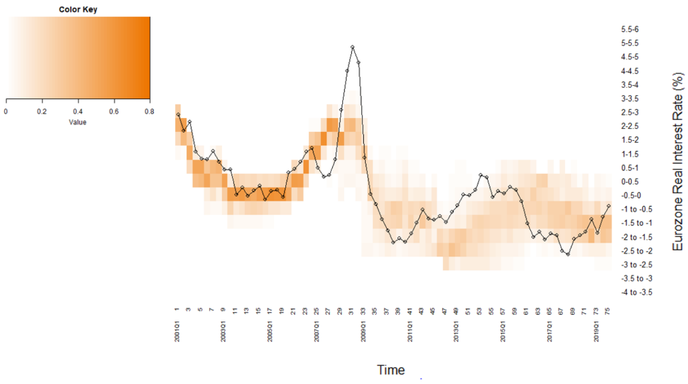


Figure 79: Best 4-Average combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1 - 2019Q3 for the BTLP ("One-Time" Estimation), displayed as heatmap.

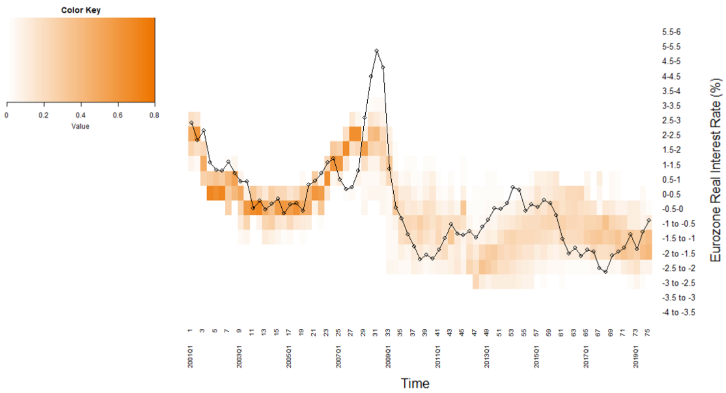


Figure 80: Best ≤4-Average combined density forecasts for (superimposed) Eurozone real interest rate 2001Q1 - 2019Q3 for the BTLP ("One-Time" Estimation), displayed as heatmap.