Exploiting Heterogeneity in the Survey of Professional Forecasters

Saerom Lee* and Tae-Hwy Lee†

April 4, 2026

Abstract

The mean response in the Survey of Professional Forecasters (SPF) is widely used to summarize individual forecasts. In this paper, we propose a novel summary forecast that enhances the predictive power of the mean response by selectively incorporating idiosyncratic signals. Our framework is motivated by the observation that while individual forecasts are highly correlated—suggesting a factor structure—they also exhibit significant heterogeneity. We treat the mean response as the primary common factor and define heterogeneity as the idiosyncratic component of each individual forecast after accounting for this commonality. Employing a factor-adjusted regularized framework, we integrate informative idiosyncratic components to improve the mean response. Using SPF data from the Federal Reserve Bank of Philadelphia and the European Central Bank, we show that incorporating these idiosyncratic components leads to significant predictive gains over the mean response.

Keywords: mean response, heterogeneity, common components, idiosyncratic components

JEL classification: C22, C32

*Department of Economics, University of California, Riverside, CA 92521. E-mail: slee839@ucr.edu.
†Department of Economics, University of California, Riverside, CA 92521. E-mail: taelee@ucr.edu.

1 Introduction

Professional forecasters and the Blue Chip Indicators serve as primary reference points for macroeconomic analysis (Chauvet and Potter (2013)). In particular, the Survey of Professional Forecasters (SPF), conducted by institutions such as the Federal Reserve Bank of Philadelphia and the European Central Bank (ECB), collects macroeconomic forecasts from a diverse group of professional forecasters. These forecasts provide valuable insights into key economic variables, such as inflation, real gross domestic product (RGDP), and the unemployment rate. Each forecaster's distinct expertise and access to private information result in heterogeneity in individual responses, a key feature this paper examines. Consequently, combining their forecasts typically produces more accurate predictions than relying on any single forecast, particularly when this heterogeneity is effectively incorporated.

However, a widely used approach to combining SPF responses is to adopt the cross-sectional mean of individual responses, which is referred to as the 'mean response'. The mean response captures the common factor underlying individual responses and is appealing due to its simplicity, as it does not require the estimation of combining weights. As noted by Genre et al. (2013), "since its inception, the forecast data collected in the SPF have normally been summarized by means of a simple average of the surveyed forecasts". Similarly, Clements et al. (2023) emphasize that "in practice, the cross-sectional median or mean ('equal weights') is routinely used".

Empirical evidence supports the competitive performance of the cross-sectional mean response. For example, Genre et al. (2013), using data from the ECB's SPF, find that only a few combination methods outperform the mean response when forecasting GDP growth and the unemployment rate. Similarly, Conflitti et al. (2015) show that the mean response remains a hard-to-beat benchmark in survey-based forecasting, largely due to the high degree of similarity among individual forecasts. More recent evidence from Elliott and Liao (2025) further demonstrates that the gains from optimally weighted combinations tend to be small, based on a long history of SPF data.

Nevertheless, the mean response fails to exploit the heterogeneity stemming from individual idiosyncratic signals, which may carry valuable information for improving forecast accuracy. As shown in Figures 1 and 2, individual forecasts tend to collectively understate or overstate target variables. Crucially, they also vary considerably, suggesting significant heterogeneity beyond the mean.

To leverage this variation, this paper proposes a novel framework for aggregating SPF responses that incorporates individual heterogeneity. We treat the mean response as the common factor and define heterogeneity through idiosyncratic components—the deviations of individual forecasts from that mean response. By selectively integrating these idiosyncratic components, we enhance the aggregate mean response and capture heterogeneity that may contain valuable information.

The empirical analysis employs data from the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters (Philadelphia-SPF) and the European Central Bank's Survey of Professional Forecasters (ECB-SPF). While forecasting performance varies across target variables and horizons, our results show that the proposed combined forecast can outperform the mean response by incorporating heterogeneity through selected idiosyncratic components—particularly in cases with larger estimation samples and shorter forecast horizons.

The main contribution of this paper is the construction of a novel summary forecast for survey data that outperforms the mean response by reflecting the heterogeneity of individual forecasts in high-dimensional settings, where the number of forecasts is large relative to the sample size. Our approach aligns with the Factor-Adjusted Regularized Model (FARM) framework introduced by Fan et al. (2020), Fan et al. (2023), and Fan et al. (2024), as we decompose individual responses into a common mean component and idiosyncratic components.

The remainder of the paper is organized as follows. Section 2 presents the methodology for combining forecasts based on FARM. Section 3 presents an out-of-sample forecast evaluation of our approach relative to the mean response using the Philadelphia-SPF and ECB-SPF. Section 4 concludes.

2 Improving the Mean Response under Heterogeneity

In this section, we propose a novel summary forecast designed to improve predictive accuracy relative to the mean response in a survey. Our approach adjusts the mean response by incorporating heterogeneity through idiosyncratic components.

To enhance the mean response, we begin by decomposing each individual forecast into a common component (the mean response) and an idiosyncratic component. Suppose there are N individual forecasts, denoted by $f_{1,t+h}, \ldots, f_{N,t+h}$, aimed at predicting the target variable y_{t+h}, where h denotes the forecast horizon. Each individual forecast $f_{i,t+h}$ for $i = 1, \ldots, N$ can be represented as follows:

$$f_{i,t+h} = \bar{f}_{t+h} + d_{i,t+h}, \quad t = 1, \ldots, T, \tag{1}$$

where $\bar{f}_{t+h} \equiv \frac{1}{N} \sum_{i=i}^{N} f_{i,t+h}$ denotes the cross-sectional mean response of individual forecasts at time t, and $d_{i,t+h} = f_{i,t+h} - \bar{f}_{t+h}$ represents the idiosyncratic response—the deviation of the individual forecast from the mean response. This idiosyncratic component, from which heterogeneity arises, reflects each forecaster's unique expertise or private information. We assume that a subset of these idiosyncratic responses contains predictive information beyond what is captured by the mean. As noted in Section 1, our approach aligns with the FARM framework, as we decompose individual responses into a common component (the mean response) and idiosyncratic components.

Our objective is to construct a combined forecast that outperforms the mean response $\bar{f}_{t+h}$ by selectively incorporating idiosyncratic components $d_{i,t+h}$ relevant to the target variable y_{t+h}.

The combined forecast, denoted by $f_{c,t+h}$, is represented as follows:

$$\begin{aligned} f_{c,t+h} &= \sum_{i=1}^{N} \beta_i f_{i,t+h} \\ &= \sum_{i=1}^{N} \beta_i \left(\bar{f}_{t+h} + d_{i,t+h} \right) \\ &= \bar{f}_{t+h} + \sum_{i=1}^{N} \beta_i d_{i,t+h}, \end{aligned} \tag{2}$$

where the coefficients β_i are combining weights that sum to one ($\sum_{i=1}^{N} \beta_i = 1$). The forecast $f_{c,t+h}$ encompasses the mean response by accounting for the heterogeneous expertise of individual forecasters, thereby enhancing predictive accuracy.

Rewriting Equation (2) in terms of forecast errors gives

$$\bar{u}_{t+h} = \sum_{i=1}^{N} \beta_i d_{i,t+h} + u_{c,t+h}, \tag{3}$$

where $\bar{u}_{t+h} \equiv y_{t+h} - \bar{f}_{t+h}$ and $u_{c,t+h} \equiv y_{t+h} - f_{c,t+h}$. The combining weights β_i are estimated by regressing the mean response forecast errors $\bar{u}_{t+h}$ on the idiosyncratic components $d_{i,t+h}$. This allows us to adjust the equal weights used in the mean response and explicitly account for forecaster heterogeneity.

We employ three penalized regression methods to estimate the combining weights β_i in high-dimensional settings, where the number of forecasts to be combined is large relative to the sample size. We first apply the Adaptive Lasso (ALasso) of Zou (2006) to select some informative idiosyncratic components. This approach operates under a sparsity assumption, positing that only a small subset of idiosyncratic responses contains predictive information for the target variable in Equation (3). The combining weights are then estimated based on the selected components, following Belloni and Chernozhukov (2013).

Second, we implement the component-wise L_2-Boosting procedure of Bühlmann (2006). The mean response is used as an initial learner in this framework. In each boosting iteration, an idiosyncratic component $d_{i,t+h}$ is selected in the direction that minimizes the mean squared forecast error (MSFE), so that the updated combined forecast encompasses the one from the previous iteration. The procedure continues until a stopping criterion is met. Following the recommendations of Hastie et al. (2009), we set the learning rate (step size) to 0.001 and the maximum number of iterations to 3,000 in our empirical applications.

Third, we consider the Ridge regression of Hoerl and Kennard (1970). While Lasso is known to struggle with 'weak' signals—predictors with negligible influence on the outcome variable—Ridge can handle them more effectively. As Shen and Xiu (2025) demonstrate, Ridge can outperform Lasso in contexts where many weak signals are present, as Lasso may fail to distinguish true signals from spurious noise. After accounting for the mean response, many individual idiosyncratic components, $d_{i,t+h}$, may represent such weak signals. In such cases, the Ridge regression provides a more robust estimation of the combining weights, β_i, in Equation (3).

Remark. Before combining SPF responses, we address the unbalanced panel structure arising from non-response and the entry or exit of forecasters in Section 3. While various imputation methods have been proposed in the survey forecasting literature—such as the EM algorithm (Capistrán and Timmermann (2009)) or autoregressive models (Genre et al. (2013))—we impute missing forecasts $f_{i,t+h}$ using the cross-sectional mean $\bar{f}_{t+h}$ of the available responses for each period. This choice is particularly advantageous for our framework, which seeks to leverage heterogeneity through idiosyncratic components. By construction, substituting a missing value with the mean response sets the corresponding idiosyncratic component to zero ($d_{i,t+h} = f_{i,t+h} - \bar{f}_{t+h} = 0$). Consequently, these imputed observations carry no idiosyncratic signal and do not distort the estimation of the combining weights β_i. This ensures that our methodology remains exclusively focused on observed idiosyncratic signals, effectively capturing the actual heterogeneity provided by forecasters while maintaining the integrity of the aggregate mean response.

3 Empirical Applications

In this section, we conduct out-of-sample forecasting exercises using two surveys of professional forecasters, the Philadelphia-SPF and ECB-SPF. These applications demonstrate that the proposed combined forecast, based on the Factor-Adjusted Regularized Model (FARM), can improve the mean response.

3.1 Heterogeneity in the Philadelphia-SPF

In this subsection, we employ the Philadelphia-SPF. The survey was launched in the fourth quarter of 1968 by the American Statistical Association and the National Bureau of Economic Research. In the second quarter of 1990, the Federal Reserve Bank of Philadelphia took over the survey. Conducted quarterly, respondents, drawn from financial institutions, banks, consulting firms, university research centers, and other private organizations, are asked to predict U.S. inflation, GDP, unemployment, interest rates, and other macroeconomic variables.

This paper focuses on quarterly forecasts for Consumer Price Index (CPI) inflation and the unemployment rate. CPI inflation is measured as the annualized quarter-over-quarter percentage change in the quarterly average of the price index. The unemployment rate shows forecasts for the quarterly average unemployment rate. We do not include RGDP in this application because the Philadelphia-SPF provides forecasts of the RGDP level rather than RGDP growth. Additional details on the survey can be found on the Federal Reserve Bank of Philadelphia's website.

We consider point forecasts for a one-quarter-ahead forecast horizon ($h = 1$), based on the most recent data available to respondents at the time of each survey. For example, although the survey was conducted in the second quarter of 2005, the most recent data available to respondents was from the first quarter of 2005. Hence, the forecast for the second quarter of 2005 is considered a one-quarter-ahead forecast.

For the out-of-sample comparison, we divide the full sample (T observations) into two sub-periods: an estimation period consisting of T_1 observations and an out-of-sample period consisting of T_2 observations. Thus, the total number of observations is $T = T_1 + T_2$. The estimation period is used to estimate the combining weights β_i in Equation (3). We then implement a rolling window procedure of size T_1 over the out-of-sample period, which comprises T_2 observations, producing T_2 one-quarter-ahead forecasts.

The sample period for CPI inflation spans from 1981Q3 to 2025Q2 ($T = 176$), while for the unemployment rate it spans from 1968Q4 to 2025Q2 ($T = 227$). We consider two sample splits: one ending in 2019Q4, and another ending in 2025Q2, each evaluated over a 10-year out-of-sample period ($T_2 = 40$). For CPI inflation, the sample splits are $(T_1, T_2) = (114, 40)$ for the sample ending in 2019Q4 and $(T_1, T_2) = (136, 40)$ for the sample ending in 2025Q2. For the unemployment rate, the corresponding splits are $(T_1, T_2) = (165, 40)$ and $(T_1, T_2) = (187, 40)$, respectively.

The raw survey initially comprises $N = 270$ forecasters for CPI inflation and $N = 462$ forecasters for the unemployment rate. Given that the number of individual forecasters varies over time due to changes in survey participation, we exclude forecasters who responded to fewer than 10% of the observations in each rolling estimation window at each time t and impute missing values with the mean response. We define N_t as the resulting number of retained forecasters in each rolling estimation window. By excluding only the extreme respondents, we are able to retain a large pool of forecasters from which to select useful idiosyncratic components.

We then estimate the combining weights β_i from Equation (3) and construct the combined forecast $f_{c,t+h}$ using Equation (2). We consider three regularized estimation methods—ALasso, L_2-Boosting, and Ridge—which we collectively refer to as the FARM methods. In addition, we consider the ordinary least squares (OLS) regression without regularization. We evaluate their performance in terms of the out-of-sample mean squared forecast error (MSFE).

Table 1 reports the MSFE of the combined forecast relative to that of the mean response from the Philadelphia-SPF. A ratio below one indicates that the combined forecast outperforms the mean response. The main findings are summarized as follows:

1. For both CPI inflation and the unemployment rate, the FARM methods significantly improve upon the mean response by leveraging forecaster heterogeneity. This is achieved by

incorporating selected idiosyncratic components, $d_{i,t+h}$, which capture unique individual signals beyond the common component.

2. This superior performance of the FARM methods holds statistically across both out-of-sample periods: the pre-COVID sample (ending in 2019Q4) and the extended sample (ending in 2025Q2). Although the p-values of the Diebold and Mariano (1995) test are at times relatively high—likely due to the limited out-of-sample size ($T_2 = 40$)—the relative MSFEs remain below one, suggesting that idiosyncratic components are a rich source of information for enhancing predictive accuracy.

3. Ridge performs well in the sample period ending in 2019Q4, whereas ALasso and L_2-Boosting—notably ALasso—show superior performance in the extended sample ending in 2025Q2. This shift suggests a structural change in the nature of forecaster heterogeneity following the COVID-19 pandemic.

4. Before the pandemic, individual forecasts tended to align closely with the mean, suggesting that idiosyncratic components, which are the source of heterogeneity, primarily consisted of many 'weak signals'. In such cases, Ridge is superior to Lasso (Shen and Xiu (2025)), explaining its strong performance in the pre-COVID sample.

5. However, the surge in economic uncertainty during the COVID-19 pandemic (Baker et al. (2020)) caused certain individual forecasts to deviate significantly from the mean. This shift resulted in more pronounced heterogeneity, where some idiosyncratic components likely transitioned into 'strong signals'. Consequently, selecting these strong signals—via ALasso or L_2-Boosting—yields better results than Ridge, which includes all idiosyncratic components, when the evaluation period includes this volatile period.

6. As expected, the OLS regression without regularization yields relative MSFEs greater than one in all cases. This underscores the inefficiency of the OLS approach when applied to a large number of forecasts without regularization.

3.2 Heterogeneity in the ECB-SPF

In this subsection, we examine the ECB-SPF. The ECB survey began in the first quarter of 1999 and has been conducted quarterly ever since. Between the first quarter of 1999 and the third quarter of 2001, the survey was conducted in the middle of each quarter; since the fourth quarter of 2001, it has been conducted in the first month of each quarter. The survey collects both point forecasts and probability distributions for three macroeconomic series: Harmonised Indices of Consumer Prices (HICP) inflation, RGDP growth, and the unemployment rate. The HICP inflation and RGDP growth are defined as year-on-year percentage changes, while the unemployment rate follows the Eurostat definition as the percentage of the labor force that is unemployed. Further details on the survey design and implementation can be found in Garcia (2003), Bowles et al. (2007), and on the ECB's website.

In this empirical application, we focus on point forecasts at the four-quarter-ahead horizon ($h = 4$), based on the most recent data available to respondents at the time of the survey. For instance, in the first quarter of 2007, respondents were asked to provide forecasts for the fourth quarter of 2007. Since the most recent data available to them was from the fourth quarter of 2006, this is considered a four-quarter-ahead forecast. In the ECB-SPF, the shortest forecast horizon is considered to be one year ahead, which corresponds to four quarters. The ECB results for $h = 4$ are therefore reported.

The sample period spans from 1999Q4 to 2025Q2 ($T = 103$). As described in Section 3.1, we consider two sample splits: one ending in 2019Q4 and another ending in 2025Q2, each evaluated over a 10-year out-of-sample period ($T_2 = 40$). For all three series, the sample splits are set to $(T_1, T_2) = (41, 40)$ for the sample ending in 2019Q4, and $(T_1, T_2) = (63, 40)$ for the sample ending in 2025Q2.

The raw dataset includes a total of $N = 114$ forecasters. We exclude forecasters who responded to fewer than 10% of the observations in each rolling estimation window and impute missing values with the mean response, following the same procedure detailed in the previous subsection. We then construct the combined forecasts, $f_{c,t+h}$, based on Equations (2) and (3), employing three regularized methods—ALasso, L_2-Boosting, and Ridge—as well as the OLS regression. We evaluate their forecasting performance in terms of MSFE.

The results are presented in Table 2. As in Section 3.1, we report the MSFE of the combined forecast relative to that of the mean response. A ratio below one indicates that the combined forecast outperforms the mean response. The main findings are summarized as follows:

1. Our approach delivers out-of-sample forecast improvements relative to the mean response for HICP inflation. However, outperforming the mean response remains challenging for RGDP growth and the unemployment rate. Compared to the results from the Philadelphia-SPF in Section 3.1, these weaker results may stem from the smaller estimation samples ($T_1 = 41$ and 63) and the longer forecast horizon ($h = 4$) in the ECB-SPF. This aligns with the findings of Aiolfi and Timmermann (2006) that the persistence of forecast performance tends to be weaker at longer horizons, suggesting that averaging across models could be more effective in improving forecasting performance in such cases.

2. Moreover, the limited out-of-sample size ($T_2 = 40$) reduces the power of the Diebold and Mariano (1995) test, often yielding high p-values even when the relative MSFE is below one. The longer forecast horizon ($h = 4$) in the ECB-SPF may further contribute to these higher p-values, compared with the shorter forecast horizon ($h = 1$) in the Philadelphia-SPF.

3. In the case of HICP inflation, ALasso and L_2-Boosting perform better in the out-of-sample period ending in 2025Q2 than in the period ending in 2019Q4. Regarding the unemployment rate, ALasso improves the mean response exclusively in the period ending in 2025Q2. Consistent with the Philadelphia-SPF results, these results likely stem from the COVID-19 pandemic, during which idiosyncratic responses deviated more sharply from the mean response. This deviation generated more pronounced heterogeneity, thereby providing stronger signals to be captured via ALasso or L_2-Boosting.

4. As expected, the OLS regression without regularization yields the relative MSFE much greater than one in all cases.

4 Conclusion

A prevailing practice in the SPF literature is to summarize individual responses by their cross-sectional mean. In this paper, we interpret this mean response as the common component and investigate idiosyncratic components to capture forecaster heterogeneity. By imputing missing values with the cross-sectional mean, we implicitly set the idiosyncratic components of missing observations to zero. We employ regularized estimation methods to incorporate heterogeneity through selected idiosyncratic components. Our empirical results show that selectively leveraging the heterogeneity embedded in idiosyncratic components can lead to significant gains in forecasting performance relative to the mean response.

References

Aiolfi, M. and Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135(1-2):31–53.

Baker, S. R., Bloom, N., Davis, S. J., and Terry, S. J. (2020). Covid-induced economic uncertainty. Technical report, National Bureau of Economic Research. No. 26983.

Belloni, A. and Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521–547.

Bowles, C., Friz, R., Genre, V., Kenny, G., Meyler, A., and Rautanen, T. (2007). The ECB survey of professional forecasters (SPF)-A review after eight years' experience. *ECB Occasional Paper*. No.59.

Bühlmann, P. (2006). Boosting for high-dimensional linear models. *Annals of Statistics*, 34:559–583.

Capistrán, C. and Timmermann, A. (2009). Forecast combination with entry and exit of experts. *Journal of Business & Economic Statistics*, 27(4):428–440.

Chauvet, M. and Potter, S. (2013). Forecasting output. *Handbook of economic forecasting*, 2:141–194.

Clements, M. P., Rich, R. W., and Tracy, J. S. (2023). Surveys of professionals. In *Handbook of Economic Expectations*, pages 71–106. Elsevier.

Conflitti, C., De Mol, C., and Giannone, D. (2015). Optimal combination of survey forecasts. *International Journal of Forecasting*, 31(4):1096–1103.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Elliott, G. and Liao, J. (2025). Combining forecasts-on why averaging beats optimal linear weights. University of California, San Diego. `https://econweb.ucsd.edu/~gelliott/Combination.pdf`.

Fan, J., Ke, Y., and Wang, K. (2020). Factor-adjusted regularized model selection. *Journal of Econometrics*, 216(1):71–85.

Fan, J., Lou, Z., and Yu, M. (2024). Are latent factor regression and sparse regression adequate?

Journal of the American Statistical Association, 119(546):1076–1088.

Fan, J., Masini, R. P., and Medeiros, M. C. (2023). Bridging factor and sparse models. *The Annals of Statistics*, 51(4):1692–1717.

Garcia, J. A. (2003). An introduction to the ECB's survey of professional forecasters. ECB Occasional Paper No. 8.

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

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The Elements of Statistical Learning: data mining, inference, and prediction*. Springer.

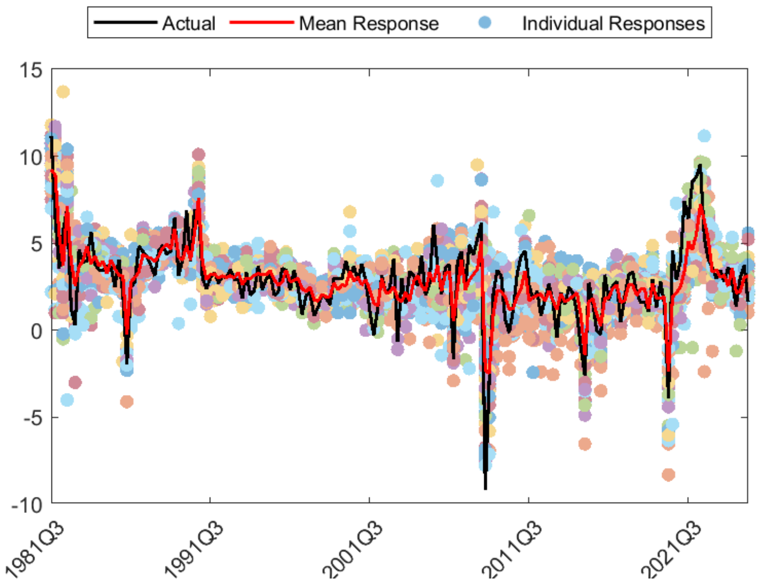
Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Shen, Z. and Xiu, D. (2025). Can machines learn weak signals? National Bureau of Economic Research Working Paper 33421.

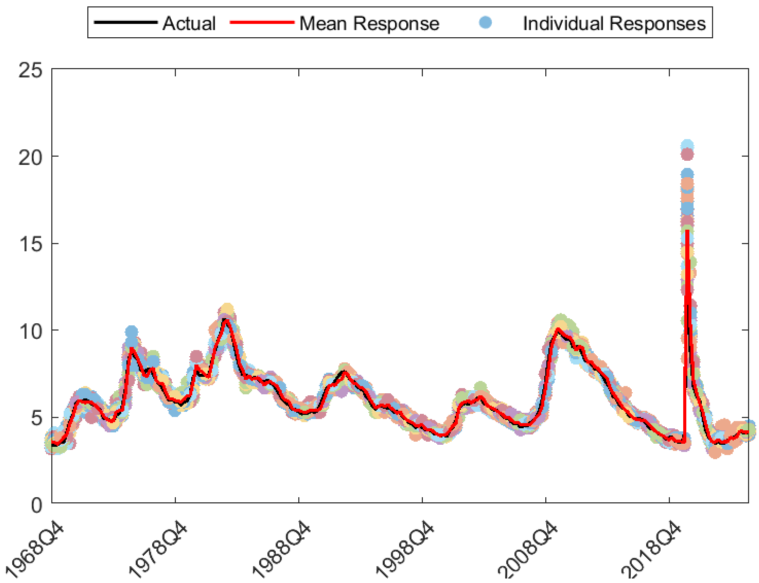
Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.

Figure 1: Philadelphia-SPF: actual series, mean response, and individual responses.

(a) CPI inflation ($h = 1$)



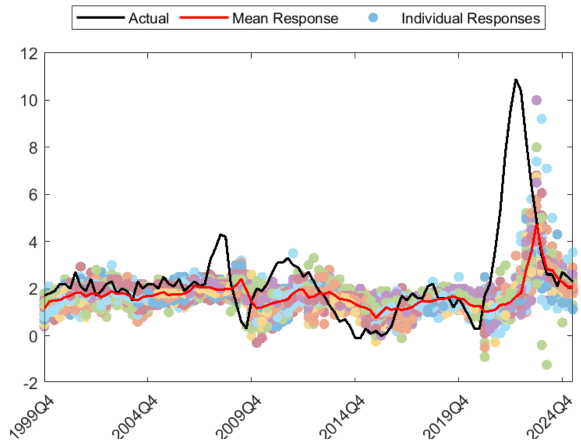
(b) Unemployment rate ($h = 1$)



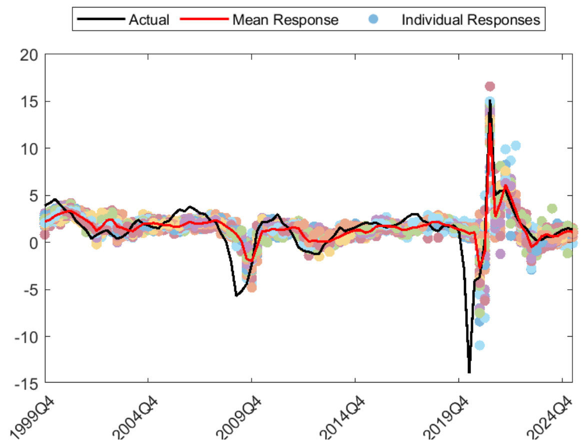
Notes: Each panel shows the actual series (black line), the mean response from the Philadelphia-SPF (red line), and individual responses (dots). The sample period for CPI inflation spans from the third quarter of 1981 to the second quarter of 2025 (1981Q3–2025Q2), while for the unemployment rate, it spans from the fourth quarter of 1968 to the second quarter of 2025 (1968Q4–2025Q2). $h = 1$ refers to one-quarter-ahead forecasts. CPI inflation denotes forecasts for the headline CPI inflation, representing annualized quarter-over-quarter percent changes in the quarterly average price index level. The unemployment rate denotes forecasts for the quarterly average unemployment rate.

Figure 2: ECB-SPF: actual series, mean response, and individual responses.

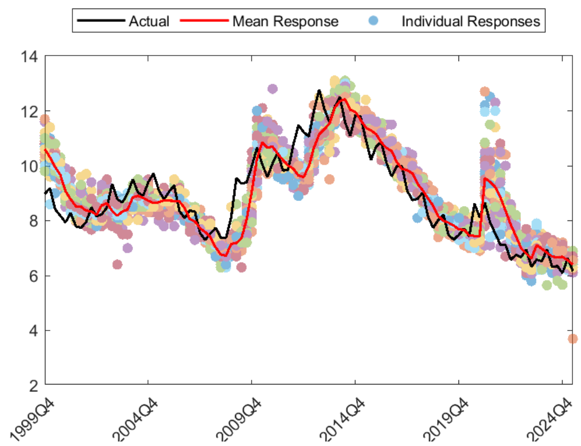
(a) HICP inflation ($h = 4$)



(b) RGDP growth ($h = 4$)



(c) Unemployment rate ($h = 4$)



Notes: Each panel shows the actual series (black line), the mean response from the ECB-SPF (red line), and individual responses (dots). The sample period spans from the fourth quarter of 1999 to the second quarter of 2025 (1999Q4–2025Q2). $h = 4$ refers to the four-quarter-ahead forecasts. HICP inflation and RGDP growth are based on year-on-year percentage changes. The unemployment rate shows the percentage of the labor force that is unemployed.

Table 1: MSFE Relative to the Mean Response: Philadelphia-SPF

Out-of-sample period	**2010Q1-2019Q4**		**2015Q3-2025Q2**	
	CPI	UNRATE	CPI	UNRATE
OLS	27.6535 (1.0000)	4.9123 (0.9786)	1.5038 (0.9983)	20.8300 (0.8570)
ALasso	1.0219 (0.5381)	1.0350 (0.5544)	0.6410 **(0.0457)**	0.7439 (0.2609)
L_2-Boosting	1.0028 (0.5731)	0.8154 **(0.0040)**	0.9943 (0.3267)	0.9621 **(0.0677)**
Ridge	0.8634 **(0.0156)**	0.7894 **(0.0027)**	0.9247 **(0.0536)**	0.9694 **(0.0654)**

Table 1 reports the MSFE ratios relative to the MSFE of the Philadelphia-SPF mean response. A ratio smaller than one indicates that the combined forecast outperforms the mean response; such cases are highlighted in blue. The p-values reported in parentheses are based on the Diebold and Mariano (1995) test. P-values smaller than 0.1 are in **bold**, indicating significance at the 10% level. CPI and UNRATE denote CPI inflation and the unemployment rate, respectively.

Table 2: MSFE Relative to the Mean Response: ECB-SPF

Out-of-sample period	**2010Q1-2019Q4**			**2015Q3-2025Q2**		
	HICP	RGDP	UNRATE	HICP	RGDP	UNRATE
OLS	8.5837 (0.9997)	12.7993 (0.9528)	12.0514 (0.9931)	17.6843 (0.9881)	834.6604 (0.9526)	28.3420 (0.9994)
ALasso	0.9588 (0.4151)	1.8948 (0.9942)	1.2930 (0.8927)	0.7677 **(0.0876)**	1.0790 (0.9618)	0.8160 (0.2247)
L_2-Boosting	1.0357 (0.5979)	1.2512 (0.9820)	1.1925 (0.8977)	0.9073 **(0.0288)**	1.0160 (0.7488)	1.2277 (0.9999)
Ridge	0.9625 (0.4005)	1.2370 (0.9781)	1.1656 (0.9083)	0.9093 **(0.0333)**	1.0159 (0.7481)	1.2175 (0.9998)

Table 2 reports the MSFE ratios relative to the MSFE of the ECB-SPF mean response. A ratio smaller than one indicates that the combined forecast outperforms the mean response; such cases are highlighted in blue. The p-values reported in parentheses are based on the Diebold and Mariano (1995) test. P-values smaller than 0.1 are in **bold**, indicating significance at the 10% level. HICP, RGDP, and UNRATE denote HICP inflation, RGDP growth, and the unemployment rate, respectively.