How to Summarize the Survey of Professional Forecasters?

Saerom Lee* and Tae-Hwy Lee†

October 28, 2025

Abstract

The mean responses in the Survey of Professional Forecasters (SPF) are commonly used to summarize individual forecasts. We propose a new summary forecast that improves the mean responses by incorporating idiosyncratic responses. We note that individual forecasts are not only highly correlated but also heterogeneous. High correlation postulates a factor structure, which we capture with the mean responses. Heterogeneity arises from idiosyncratic responses after accounting for the common mean responses. Using the SPFs from the Philadelphia Fed and European Central Bank, we show that incorporating informative idiosyncratic responses can lead to improved forecasting performance over the mean responses.

Keywords: mean responses, common components, idiosyncratic components, high dimension, sparsity, imputing missing responses

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

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 the idiosyncratic component of this heterogeneity is effectively incorporated.

However, a widely used approach to combining the SPF responses is to summarize them using the cross-sectional mean responses. The mean responses capture the common factor underlying individual responses and are appealing due to their simplicity, as they do 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 responses. For example, Genre et al. (2013), using data from the ECB's SPF, find that only a few combination methods outperform the mean responses 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, while the mean responses capture the common component across forecasts, they fail to exploit the individual forecasts' heterogeneity from idiosyncratic signals, which may carry valuable information to improve forecast accuracy. As shown in Figures 1 and

2, individual forecasts tend to jointly understate or overstate the target variables. Crucially, they also vary considerably, which suggests the presence of important idiosyncratic responses deviating from the mean responses. We consider the mean response as the common factor of individual responses. Some of the idiosyncratic responses (differences between individual forecasts and the mean response) may be useful, and incorporating them can improve the mean response.

Another challenge in combining the SPF responses arises from missing data, which occurs as forecasters enter and exit the survey panel over time. This dynamic structure leads to an unbalanced panel, complicating the estimation and comparison of combined forecasts. Effectively addressing this issue requires a method that can handle missing values in a robust way that does not affect forecast combinations.

This paper proposes a new approach for summarizing the SPF responses that addresses the two aforementioned challenges. First, we enhance the mean responses by selectively incorporating idiosyncratic components that constitute the observed heterogeneity. Second, we impute missing responses using the mean responses, which, by construction, sets the corresponding idiosyncratic components to zero. Therefore, this imputation does not affect the estimation of combining weights.

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). Although forecasting performance varies across target variables and forecast horizons, our results show that the proposed combined forecast can improve the mean responses by employing selected idiosyncratic components.

The main contribution of this paper is to improve upon the mean responses by selectively incorporating idiosyncratic components in high-dimensional settings where the number of forecasts to be combined is large relative to the sample size. Our approach aligns with the Factor-Adjusted Regularization Models (FARM) introduced by Fan et al. (2020), Fan et al. (2023), and Fan et al. (2024), as we decompose the individual responses into the common response (the mean response) and idiosyncratic responses.

The remainder of the paper is organized as follows. Section 2 presents the methodology for combining forecasts using FARM, incorporating both the common factor and idiosyncratic components. Section 3 presents an out-of-sample forecast evaluation of our approach relative to the mean responses using the Philadelphia-SPF and ECB-SPF. Section 4 concludes.

2 How to Improve the Mean Responses in Summarizing the SPFs?

In this section, we propose a new summary forecast designed to improve predictive accuracy relative to the mean responses in a survey. Our approach adjusts the mean responses by incorporating some informative idiosyncratic components in two steps: first, we impute missing responses using the mean responses (as discussed in Section 2.1), and second, we selectively incorporate informative idiosyncratic responses (as discussed in Section 2.2).

2.1 Handling Missing Values

Missing values refer to the absence of observations at certain points in time. In a survey, these missing values primarily result from two sources: non-response and the entry or exit of forecasters. Non-response occurs when a forecaster does not provide a forecast. In addition, entry and exit dynamics introduce further missing observations. Newly entering forecasters do not have responses for earlier periods, while exiting forecasters stop providing responses in subsequent periods.

Various methods have been proposed to handle missing values, ranging from simple approaches, such as mean or mode imputation, to more advanced techniques using machine learning. For example, Capistrán and Timmermann (2009) use an EM algorithm to backfill missing observations, Genre et al. (2013) employ autoregressive models, and Lee and Seregina (2025) apply random forests. A comprehensive review of imputation methods is provided by Lin and Tsai (2020).

In this paper, we impute missing values using the cross-sectional mean responses of the available forecasts at each period. This approach preserves the mean responses, as the imputed mean responses remain unchanged. For any missing observation, the corresponding idiosyncratic component is set to zero.

2.2 Incorporating Idiosyncratic Components

To improve the mean responses, we start by decomposing each individual forecast into a common component (mean response) and an idiosyncratic component. Suppose there are N individual forecasts, denoted by $f_{1,t+h}, \ldots, f_{N,t+h}$, each 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:

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

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 $d_{i,t+h} = f_{i,t+h} - \bar{f}_{t+h}$ is the idiosyncratic response (the deviation of the individual response from the mean response). This idiosyncratic response, from which heterogeneity arises, reflects each forecaster's unique expertise or private information. We assume that some idiosyncratic responses, likely a small number, contain predictive information beyond what is reflected in the mean response. As mentioned in Section 1, our approach aligns with the FARM, as we decompose the individual responses into the common response (the mean response) and idiosyncratic responses.

When a missing response $f_{i,t+h}$ is imputed using the mean response $\bar{f}_{t+h}$, the corresponding idiosyncratic component is zero ($d_{i,t+h} = f_{i,t+h} - \bar{f}_{t+h} = \bar{f}_{t+h} - \bar{f}_{t+h} = 0$). Accordingly, the proposed combined forecast in Equations (2) and (3) below remains unaffected by this mean-imputation.

Our objective is to construct a combined forecast of individual responses in a survey that outperforms the mean responses $\bar{f}_{t+h}$ by incorporating some idiosyncratic components $d_{i,t+h}$ that contain predictive information relevant to the forecast target variable y_{t+h}. The combined forecast is written as

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

where the coefficients β_i are the combining weights, which sum to one $\left(\sum_{i=1}^{N} \beta_i = 1\right)$. The combined forecast $f_{c,t+h}$ encompasses the mean responses by selectively incorporating idiosyncratic responses that enhance 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 can be 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 in the mean responses, thereby accounting for the heterogeneous expertise of individual forecasters.

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. First, we apply the Adaptive Lasso of Zou (2006) to select informative idiosyncratic components, under the sparsity assumption that only a subset of idiosyncratic responses contains predictive information for the target variables in Equation (3). The combining weights are then estimated based on the selected components, following Belloni and Chernozhukov (2013), and this approach is referred to as Post-ALasso in Section 3.

Second, we implement the component-wise L_2-Boosting procedure of Bühlmann (2006). The mean response is used as an initial learner in our combining procedure. In each boosting iteration, an idiosyncratic component $d_{i,t+h}$ is selected in the direction that minimizes the mean squared forecast errors, so that the updated combined forecast encompasses the one from the previous iteration. The procedure continues until a stopping criterion is met. The L_2-Boosting algorithm involves two hyperparameters: the step size (learning rate) and the maximum number of iterations. We set the step size to 0.001 and the maximum number of iterations to 3000 in our empirical applications, following the recommendations of Hastie et al. (2009).

Third, we consider the Ridge regression of Hoerl and Kennard (1970). While Lasso is known to struggle with "weak" signals (predictors that have negligible influence on the outcome variable), Ridge can handle them more effectively. Shen and Xiu (2025) theoretically demonstrate that Ridge outperforms Lasso when many weak signals are present due to Lasso's learning limitations in distinguishing true signals from spurious ones in such contexts. After accounting for the mean responses, many individual responses may be only weakly correlated with the target variable; in other words, many idiosyncratic components $d_{i,t+h}$ may be weak signals. In such cases, Ridge provides a more reliable estimate of the combining weights β_i in Equation (3).

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 responses.

3.1 How to Combine the Individual Responses in 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 were 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 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-step-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 is time-varying due to changes in survey participation, we exclude forecasters who responded less than

10% of the observations in each rolling estimation window at each time t and impute missing values with the mean responses. 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 (Post-ALasso, L_2-Boosting, and Ridge), which we collectively refer to as the FARM method. Additionally, we consider the standard ordinary least squares (OLS) regression. We evaluate their performance in terms of the out-of-sample mean squared forecast errors (MSFE).

Table 1 reports the MSFE ratios relative to the MSFE of the mean responses of the Philadelphia-SPF. The main findings are summarized as follows:

1. For both CPI inflation and the unemployment rate targets, the FARM methods significantly improve upon the mean responses by incorporating selected idiosyncratic components $d_{i,t+h}$. According to the Diebold and Mariano (1995) test, some of the FARM methods statistically outperform the mean responses in both out-of-sample periods: the pre-COVID sample (ending in 2019Q4) and the extended sample (ending in 2025Q2). In our applications, the Diebold and Mariano (1995) test's p-values are often large even though the relative MSFEs are less than one, a likely result of the small out-of-sample size ($T_2 = 40$). These results consistently suggest that the idiosyncratic components contain valuable predictive information for the target variables.

2. Ridge performs well in the sample period ending in 2019Q4, whereas Post-ALasso performs better in the extended sample period ending in 2025Q2. This result may stem from a change in the predictive value of individual responses before and after the COVID-19 pandemic. Before the pandemic, individual responses likely aligned more closely with the mean responses. As a result, the idiosyncratic components are likely "weak signals" in the sample ending in 2019Q4. Consistent with theoretical findings (Shen and Xiu (2025)), Ridge, which is superior to the Lasso in the presence of many weak signals, performs well. However, when we include the COVID-19 period in the out-of-sample period (ending in

2025Q2), economic uncertainty increases substantially (Baker et al. (2020)). Including this volatile period in the out-of-sample evaluation, there may be some individual forecasts that deviate significantly from the mean response. Consequently, some idiosyncratic components likely become strong signals that are effectively selected by the Post-ALasso procedure.

3. As expected, the OLS regression without regularization yields relative MSFE 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 How to Combine the Individual Responses in 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 less than 10% of the observations in each rolling estimation window and impute missing values with the mean responses, 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), using the three regularized estimation methods (Post-ALasso, L_2-Boosting, and Ridge) and the OLS regression. We evaluate their forecast performance in terms of MSFE.

The results are presented in Table 2. The main findings are summarized as follows:

1. Our approach delivers out-of-sample forecast improvements for HICP inflation and the unemployment rate in the ECB-SPF. However, the results are less statistically significant than those obtained for the Philadelphia-SPF in the previous subsection. This weaker significance likely reflects the smaller estimation samples ($T_1 = 41$ or 63) and the longer forecast horizon ($h = 4$) in the ECB-SPF. Moreover, the limited out-of-sample size ($T_2 = 40$) reduces the power of the Diebold and Mariano (1995) test, often yielding large 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 larger p-values compared with the Philadelphia-SPF results at $h = 1$.

2. Similar to the Philadelphia-SPF results, Post-ALasso performs better in the out-of-sample period ending in 2025Q2 than in the out-of-sample period ending in 2019Q4. The reason for this improvement may be that many idiosyncratic responses deviated more from the mean responses during the COVID-19 pandemic, generating stronger signals for the Lasso to capture.

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

4 Conclusion

A common practice when using the SPFs is to take the mean of individual responses. We interpret this mean response as the common component shared across individual respondents and investigate the idiosyncratic responses for potential gains in forecast accuracy. Typically, the SPF includes a large number of individual forecasts at each point in time. By decomposing these into the mean response and deviations from it, we assume that a small number of idiosyncratic responses (i.e., deviations from the mean response) may contain useful predictive information beyond what is captured by the mean response. Under the sparsity assumption, we apply a regularized estimation of our model in Equation (3). Our empirical results demonstrate that leveraging the additional predictive content in idiosyncratic responses can lead to improved forecasting performance over the mean responses.

References

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.

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., Friedman, J. H., 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.

Lee, T.-H. and Seregina, E. (2025). Combining forecasts under structural breaks using graphical lasso. *International Journal of Forecasting*. ISSN: 0169–2070.

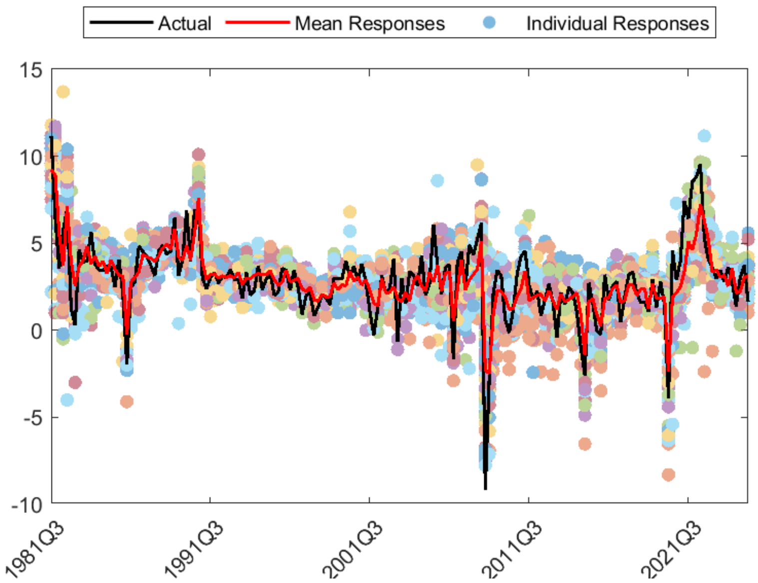
Lin, W.-C. and Tsai, C.-F. (2020). Missing value imputation: a review and analysis of the literature (2006–2017). *Artificial Intelligence Review*, 53(2):1487–1509.

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

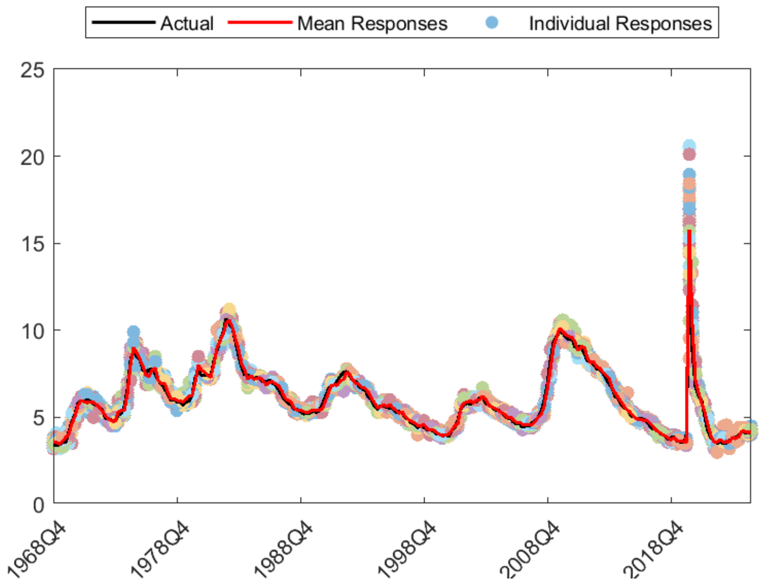
Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.

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

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



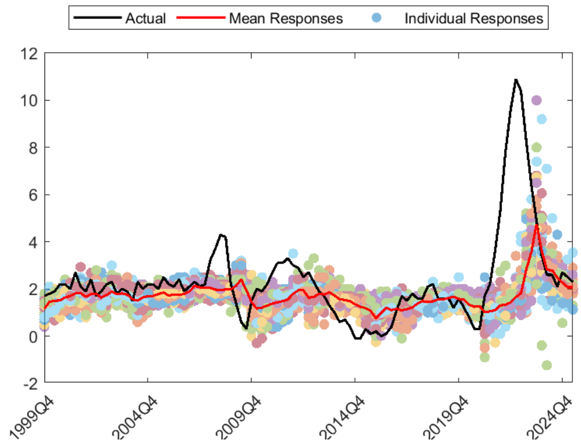
(b) Unemployment rate ($h = 1$)



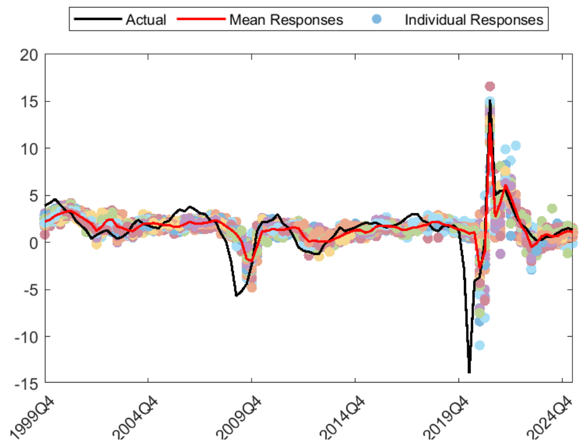
Notes: Each panel shows the actual series (black line), the mean responses from the Philadelphia-SPF (red line), and individual responses (dots). The sample period for CPI inflation spans from the third quarter of 1981 to the second quarter of 2025 (1981Q3–2025Q2), while for the unemployment rate, it 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 responses, and individual responses.

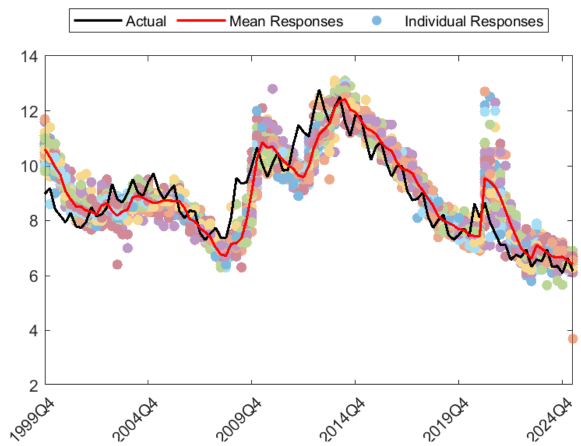
(a) HICP inflation ($h = 4$)



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



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



Notes: Each panel shows the actual series (black line), the mean responses 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 Responses: 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)
Post-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 mean responses of the Philadelphia-SPF. A relative MSFE ratio smaller than one indicates that the combined forecast from a method outperforms the mean responses, and it is highlighted in blue. The p-values reported in parentheses are based on the Diebold and Mariano (1995) test. Those p-values smaller than 0.1 are in **bold** indicating significance at the 10% level. CPI refers to CPI inflation, and UNRATE refers to the unemployment rate.

Table 2: MSFE Relative to the Mean Responses: 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)
Post-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 mean responses of the ECB-SPF. A relative MSFE ratio smaller than one indicates that the combined forecast from a method outperforms the mean responses, and it is highlighted in blue. The p-values reported in parentheses are based on the Diebold and Mariano (1995) test. Those p-values smaller than 0.1 are in **bold** indicating significance at the 10% level. HICP refers to HICP inflation, RGDP to RGDP growth, and UNRATE to the unemployment rate.