On the Wisdom of Crowds (of American and European Economists)

Francis X. Diebold
University of Pennsylvania
and NBER

Aarón Mora
University of South Carolina

Minchul Shin
Federal Reserve Bank of Philadelphia

October 29, 2024

Abstract: We study the properties of macroeconomic survey forecast response averages as the number of survey respondents grows. Such averages are "portfolios" of forecasts. We characterize the speed and pattern of the gains from diversification and their eventual decrease with portfolio size (the number of survey respondents) in both (1) the key real-world data-based environments of U.S. and European Surveys of Professional Forecasters (SPFs), and (2) the theoretical model-based environment of equicorrelated forecast errors. We proceed by proposing and comparing various data-based and model-based "crowd size signature plots", which summarize the forecasting performance of k-average forecasts as a function of k, where k is the number of forecasts in the average. We then estimate the equicorrelation model for growth and inflation SPFs by choosing model parameters to minimize the divergence between data and model signature plots. The results indicate near-perfect fit of the equicorrelation model for both growth and inflation, for both the U.S. and European SPFs, with much greater gains from diversification for inflation forecasts than for growth forecasts, and with the gains decreasing quite quickly, so that fewer SPF respondents than currently used may be adequate.

Acknowledgments: For helpful discussions, we thank participants at the International Conference on Macroeconomic Analysis and International Finance (Rethimno, Crete, 2024), and the Conference on Real-Time Data Analysis, Methods and Applications in Macroeconomics and Finance (Bank of Canada, Ottawa, 2024).

Key words: Macroeconomic surveys of professional forecasters, forecast combination, model averaging, equicorrelation

JEL codes: C5, C8, E3, E6

Contact: fdiebold@sas.upenn.edu, aaron.mora@moore.sc.edu, minchul.shin@phil.frb.org

Contents

1 Introduction and Basic Framework **1**

2 Data-Based Crowd Size Signature Plots **3**

2.1 The U.S. Survey of Professional Forecasters 4

2.2 Estimated Crowd Size Signature Plots 7

3 Model-Based Crowd Size Signature Plots **13**

3.1 The Equicorrelation Model 13

3.2 Analytic Crowd Size Signature Plots 17

4 Matching Data-Based and Model-Based Signature Plots **20**

5 Comparative Analysis of the U.S. and European SPFs **24**

6 Conclusions and Directions for Future Research **30**

References **33**

Appendices **38**

A Data Definitions and Sources **38**

B Generalized Equicorrelation **39**

C Properties of the Matching Estimator **41**

D Testing for Equicorrelation **45**

1 Introduction and Basic Framework

The wisdom of crowds, or lack thereof, is traditionally and presently a central issue in psychology, history, and political science; see for example Surowiecki (2005) regarding wisdom, and Aliber et al. (2023) regarding lack thereof. Perhaps most prominently, however, the wisdom of crowds is also—and again, traditionally and presently—a central issue in economics and finance, where heterogeneous information and expectations formation take center stage.

In this paper we focus on economics and finance, studying the "wisdom" of "crowds" of professional economists. We focus on the U.S. and European Surveys of Professional Forecasters (SPFs), which are important not only in facilitating empirical academic research in macroeconomics and financial economics, but also—and crucially—in guiding real-time policy, business, and investment management decisions.[1,2]

In particular, we study SPF crowd behavior as crowd size grows, asking precisely the same sorts of questions of SPF "forecast portfolios" that one asks of financial asset portfolios: How quickly, and with what patterns, do diversification benefits become operative, and eventually dissipate, as portfolio size (the number of SPF forecasters) grows, and why? Do the results differ across SPF variables (growth vs. inflation) and/or continents (U.S. vs. Europe), and if so, why? What are the implications for SPF size and design?

We answer the above questions by introducing and using "crowd size signature plots", which summarize the forecasting performance of k-average forecasts as a function of k, where k is the number of forecasts in the average. We examine not only data-based SPF signature plots (empirically), but also model-based signature plots (analytically, based on a simple model of forecast-error equicorrelation), after which we proceed to estimate the equicorrelation model by minimizing divergence between the data-based model-based signature plots.

To fix ideas and notation, let us sketch the basic framework with some precision. Let N refer to a set of forecasts with $N \times 1$ zero-mean time-t error vector e_t, $t = 1, ..., T$, and let $k \leq N$ refer to a subset of forecasts. We consider k-forecast averages, and we seek to characterize k-forecast mean-squared forecast error (MSE). For a particular k-forecast average corresponding to group g_k^*, the forecast error is just the average of the individual

[1] On real-time policy and its evaluation, see John Tayor's inaugural NBER Feldstein Lecture at `https://www.hoover.org/sites/default/files/gmwg-empirically-evaluating-economic-policy-in-real-time.pdf`.

[2] For an introduction to the U.S. SPF, see the materials at `https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/survey-of-professional-forecasters`. For an introduction to the European SPF. see the materials at `https://www.ecb.europa.eu/stats/ecb_surveys/survey_of_professional_forecasters/html/index.en.html`.

forecast errors, so we have

$$MSE_T^*(k) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{k}\sum_{i\in g_k^*} e_{it}\right)^2. \tag{1}$$

For any choice of k, however, there are $\binom{N}{k}$ possible k-forecast averages. We focus on the k-average $MSE_T^*(k)$ given by equation (1), averaged across all groups of size k,

$$MSE_{NT}^{avg}(k) = \frac{1}{\binom{N}{k}}\sum_{g_k=1}^{\binom{N}{k}}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{k}\sum_{i\in g_k} e_{it}\right)^2\right), \tag{2}$$

where g_k is an arbitrary member of the set of groups of size k.

Among other things, we are interested in:

(a) Tracking and visualizing $MSE_{NT}^{avg}(k)$ as k grows ("crowd size $MSE_{NT}^{avg}(k)$ signature plots");

(b) Tracking and visualizing the *change* (hopefully improvement) in $MSE_{NT}^{avg}(k)$ from adding one more forecast to the pool (i.e., moving from k to $k+1$ forecasts),

$$DMSE_{NT}^{avg}(k) = MSE_{NT}^{avg}(k) - MSE_{NT}^{avg}(k+1), \tag{3}$$

as k grows ("crowd size $DMSE_{NT}^{avg}(k)$ signature plots");

(c) Tracking and visualizing the average, best, and worst $MSE_{NT}^{avg}(k)$ performances from k-averaging relative to no averaging,

$$R_{NT}^{avg}(k) = \frac{MSE_{NT}^{avg}(k)}{MSE_{NT}^{avg}(1)} \tag{4}$$

$$R_{NT}^{max}(k) = \frac{MSE_{NT}^{max}(k)}{MSE_{NT}^{avg}(1)} \tag{5}$$

$$R_{NT}^{min}(k) = \frac{MSE_{NT}^{min}(k)}{MSE_{NT}^{avg}(1)}, \tag{6}$$

as k grows ("crowd size $R_{NT}^{avg}(k)$ signature plots");

(d) Tracking and visualizing not just *mean* squared-error performance as k grows, as in all of the above signature plots, but rather the complete *distributional* squared-error performance as k grows ("crowd size $f_{NT}^{avg}(k)$ signature plots", where $f_{NT}^{avg}(k)$ is the average empirical *distribution* of e_{it}^2 for crowd size k);

(e) Understanding paths and patterns in the above signature plots as k grows, in both leading real-world data-based environments (U.S. and European SPFs), and in leading stylized model-based environments (equicorrelation);

(f) Using SPFs to estimate and assess equicorrelation models;

(g) Understanding similarities and differences in results across variables (growth vs. inflation) and economic regions (U.S. vs. Eurozone);

(h) Drawing practical implications for SPF design.

We proceed as follows. In section 2 we study the U.S. SPF, and we estimate its crowd size signature plots. In section 3 we study an equicorrelation model, and we characterize its crowd size signature plots analytically for any parameter configuration. In section 4 we estimate and assess the equicorrelation model by choosing its parameters to minimize the divergence between model (equicorrelation) and data (SPF) signature plots. In section 5 we provide a comparative analysis of the European SPF. In section 6 we conclude and sketch several directions for future research.

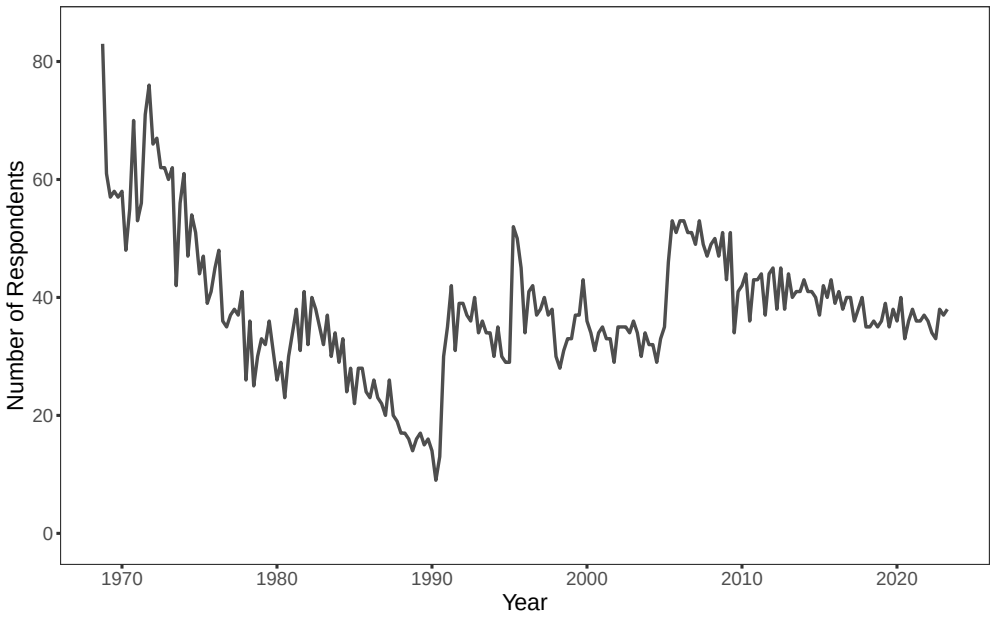
2 Data-Based Crowd Size Signature Plots

The Survey of Professional Forecasters is a quarterly survey covering several U.S. macroeconomic variables. It was started in 1968Q4 and is currently conducted and maintained by the Federal Reserve Bank of Philadelphia.[3] In Figure 1 we show the evolution of the number of forecast participants, which declined until 1990Q2, when the Federal Reserve Bank of Philadelphia took control of the survey, after which it has had approximately 40 participants. Participants stayed for 15 quarters on average, with a minimum of 1 quarter and a maximum of 125 quarters.

We will analyze SPF point forecasts for real output growth ("growth") and GDP deflator inflation ("inflation"), for forecast horizons $h = 1, 2, 3,$ and 4, corresponding to short-,

[3]For a recent introduction see Croushore and Stark (2019).

Figure 1: SPF Participation



Notes: We show the number of participants in the U.S. Survey of Professional Forecasters from 1968Q4 to 2023Q2.

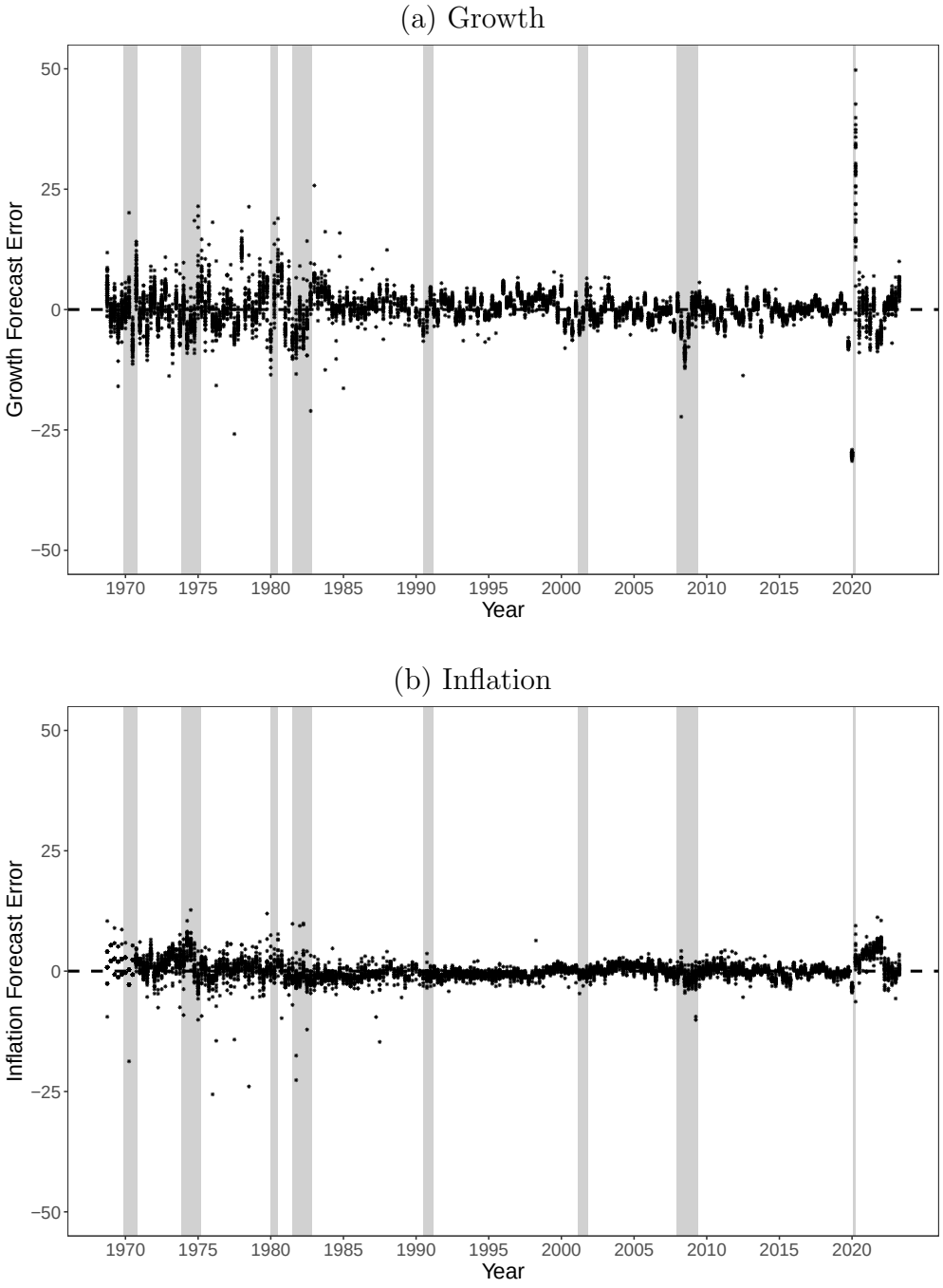
medium-, and longer-term forecasts.[4] Our sample period is 1968Q4-2023Q2, during which the survey panel had 38 participants per survey on average.

2.1 The U.S. Survey of Professional Forecasters

Because the individual forecast errors drive everything as per equation (2), it is worth taking some time to examine their evolving cross-sectional distributions, period-by-period. In Figure 2 we show the cross-section of 1-step-ahead forecast errors for growth and inflation (with every forecast error plotted vertically for each period), and in Figure 3 we show the corresponding time-series of cross-sectional means and standard deviations. Several features are apparent:

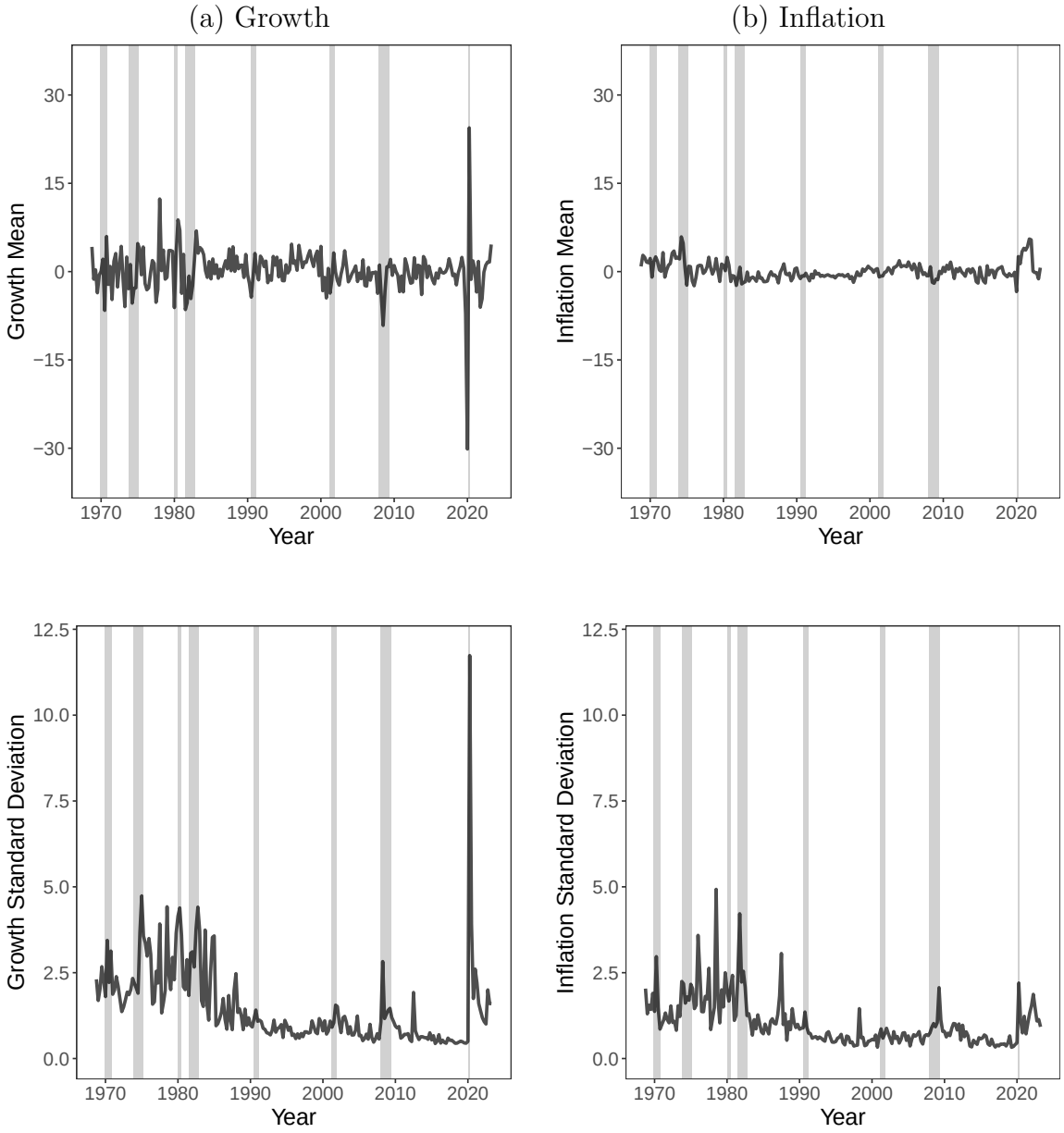
[4]The SPF contains quarterly level forecasts of real GDP and the GDP implicit price deflator. We transform the level forecasts into growth and inflation forecasts by computing annualized quarter-on-quarter growth rates, and we compute the corresponding forecast errors using realized values as of December 2023. We take the realized values from the "Forecast Error Statistics for the SPF" report of the Federal Reserve Bank of Philadelphia, which compiles historical values of various variables in the SPF. For details, see Appendix A and `https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/error-statistics`.

Figure 2: SPF Individual Forecast Errors, 1968Q4-2023Q2



Notes: We show individual 1-step-ahead forecast errors for growth and inflation. (For each quarter we show the set of all individual forecast errors, stacked vertically.) Gray shaded regions denote NBER-designated recessions.

Figure 3: SPF Forecast Error Means and Standard Deviations, 1968Q4-2023Q2



Notes: We show time series of cross-sectional means and standard deviations of individual 1-step-ahead forecast errors for growth and inflation. Gray shaded regions denote NBER-designated recessions.

1. The Great Moderation is clearly reflected in both the growth and inflation error distributions, which have noticeably reduced variability from the end of the Volcker Recession to the start of the Great Recession.

2. Growth tends to be over-predicted when entering recessions; that is, the mean of the growth error (actual minus predicted) distribution tends to be negative. Hence recessions tend to catch forecasters by surprise. The Pandemic Recession provides the most extreme example, as the mean growth error plunges.

3. Growth is, however, sometimes systematically *under*-predicted during brisk recoveries. The Pandemic Recession again provides the most extreme example, as the mean growth error leaps skyward.

4. Inflation shows little such systematic over- or under-prediction when entering recessions, with the exception of the Oil Shock Recession, when inflation was noticeable under-predicted.

5. The variability of the output error distribution tends to increase during recessions, particularly during the two most recent recessions (the Great Recession and the Pandemic Recession), reflecting greater disagreement among forecasters. The same is true of the inflation error distribution during those two recessions.

6. The unusual behavior of the inflation error distribution following the Pandemic Recession is clearly revealed. The mean error is always positive there (i.e., forecasters tended to under-predict), with the amount of under-prediction first growing and then shrinking, and the error variability follows a similar path.

2.2 Estimated Crowd Size Signature Plots

In principle we want to compute $MSE_{NT}^{avg}(k)$ as per equation (2), but that is impossible in practice unless N is very small, due to the potentially huge number of different k-average forecasts.[5] Hence we proceed by approximating $MSE_{NT}^{avg}(k)$ as follows:

1. Randomly select a k-average forecast g_k^*, and calculate $MSE_T^*(k)$ as per equation (1),

[5]For example, with $N = 40$ and $k = 20$, which are realistic values for the SPF and other surveys of forecasters, we obtain $\binom{N}{k} = 1.4 \times 10^{11}$.