JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **26**: 1059–1078 (2011)
Published online 8 September 2010 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.1206

ASSESSING THE TEMPORAL VARIATION OF MACROECONOMIC FORECASTS BY A PANEL OF CHANGING COMPOSITION

JOSEPH ENGELBERG,[a]* CHARLES F. MANSKI[b] AND JARED WILLIAMS[c]

[a] *Department of Finance, Kenan-Flagler Business School, University of North Carolina, Chapel Hill, NC, USA*
[b] *Department of Economics and Institute for Policy Research, Northwestern University, Evanston, IL, USA*
[c] *Department of Finance, Smeal College of Business, Penn State University, University Park, PA, USA*

SUMMARY

This paper calls attention to the problem of changing panel composition in surveys of forecasters and documents the problem in the Survey of Professional Forecasters. To study the temporal variation of forecasts, we recommend analysis of the time series of predictions made by individual forecasters. This makes transparent the heterogeneity of the panel and avoids improper inferences due to changing panel composition. In the absence of knowledge of the process determining panel composition, we warn against the traditional practice of aggregate time series analysis, which conflates changes in the expectations of individual forecasters with changes in the composition of the panel. Should analysis of aggregated predictions be thought desirable as a simplifying device, we recommend analyses of sub-panels of fixed composition.
Copyright © 2010 John Wiley & Sons, Ltd.

Received 22 January 2009

1. INTRODUCTION

A number of surveys periodically report the macroeconomic predictions of panels of professional forecasters. Perhaps best known are the venerable Livingston Survey and the Survey of Professional Forecasters, begun in 1946 and 1968 respectively, both presently conducted by the Federal Reserve Bank of Philadelphia. Others are the Bank of England's Survey of External Forecasters, the CESifo World Economic Survey, the INSEE's Monthly Business Survey, The Duke/CFO Magazine Global Business Outlook Survey, and a panel of the National Association of Business Economics.

To study the temporal variation of forecasts, it is common to aggregate the predictions reported by panel members at each administration of the survey and analyze the time series of the aggregated predictions. See, for example, Hafer and Hein (1985), Fair and Shiller (1989), Pennacchi (1991), Baghestani (1994, 2006), Thomas (1999), Romer and Romer (2000), Ball and Croushore (2003), and Campbell (2007). Summary reports of survey findings traditionally take this form. Consider the quarterly Survey of Professional Forecasters (SPF). In February 2008, the Philadelphia Fed issued a release of findings from the survey administered in the first quarter of 2008, with this opening statement: 'The outlook for growth in the first half of 2008 looks much weaker now than it did three months ago, according to 50 forecasters surveyed by the Federal Reserve Bank of Philadelphia. . . . Growth in the current quarter is projected at an annual rate of 0.7 percent, down from the projection of 2.2 percent in last year's fourth-quarter survey.'[1]

* Correspondence to: Joseph Engelberg, Department of Finance, University of North Carolina Kenan-Flagler Business School, McColl Building CB 3490, Chapel Hill, NC 27 599-3490, USA. E-mail: joseph_engelberg@unc.edu
[1] See http://www.philadelphiafed.org/files/spf/survq108.html.

Copyright © 2010 John Wiley & Sons, Ltd.

Interpretation of the temporal variation in an aggregated prediction can be problematic when forecasters are heterogeneous, and the interpretative problem is exacerbated when panel composition changes over time. First consider heterogeneity with a panel of fixed composition. When the Philadelphia Fed reported that growth is projected at an annual rate of 0.7%, one cannot know whether this was a consensus across the 50 forecasters or whether they disagreed sharply in their predictions. Nor can one know whether all panel members revised their beliefs downward between the fourth quarter of 2007 (4Q2007) and the first quarter of 2008 (1Q2008). This and other difficulties of inference stemming from forecaster heterogeneity have long been recognized by researchers. See, for example, Zarnowitz and Lambros (1987), Keane and Runkle (1990), Giordani and Soderlind (2003), Pesaran and Weale (2006, ch. 5), Patton and Timmerman (2008), and Engelberg *et al.* (2009).

Changing panel composition has not received similar attention. Although the Philadelphia Fed release of findings stated that 50 forecasters participated in the survey, this actually was the number of participants in the 1Q2008 survey. The number of participants in the 4Q2007 survey was 48, of whom only 42 participated in the 1Q2008 survey.[2] Thus 14 forecasters participated in only one of the two surveys, six participating only in 4Q2007 and eight only in 1Q2008. To an unknown extent, the dramatic weakening in beliefs about future growth reported in the release of findings could be an artifact of changing panel composition. At the extreme, six optimistic forecasters may have been replaced by eight pessimistic ones.

Although the Philadelphia Fed release quoted above does not mention changing panel composition as a possible source of temporal variation in the aggregated forecast, Fed releases have occasionally remarked on this possibility. For example, a release in early 2007 observed that new panel entrants had lower inflation expectations than those who had participated in a previous survey. The release stated: 'This suggests that a changing composition of the panel of forecasters over the last two surveys also contributes to the downward revision to the consensus long-term CPI inflation outlook.'[3]

To quantify the possible consequences of changing panel composition in a simple setting, suppose that four forecasters labeled (A, B, C, D) are surveyed in the first quarter of year t and report heterogeneous subjective probability distributions for inflation in year $t+1$. In particular, their subjective medians (M) and interquartile ranges (IQR) are [M(A) = 0.04, IQR(A) = 0.02], [M(B) = 0.04, IQR(B) = 0.01], [M(C) = 0.02, IQR(C) = 0.02], and [M(D) = 0.02, IQR(D) = 0.01]. Suppose further that a hypothetical 'representative forecaster' is defined to have the average of the reported values of M and IQR. Thus the representative forecaster, labeled R, has [M(R) = 0.03, IQR(R) = 0.015]. Now move ahead to the second quarter of year t and let the panel be asked again for their expectations about inflation in year $t+1$. Suppose that all four forecasters continue to have the same expectations that they reported in the first quarter, but only three of them respond to the second survey. If forecaster D does not respond, the representative forecast in the second quarter is [M(R) = 0.033, IQR(R) = 0.017]. Hence the temporal variation in the representative forecast makes it appear that forecasters have become more pessimistic and more uncertain about inflation. On the other hand, if forecaster A does not respond to the second survey, the representative forecast becomes [M(R) = 0.027, IQR(R) = 0.013]. In this case, the temporal variation in the representative forecasts makes it appear that forecasters have become more optimistic and more certain about inflation.

[2] Here 'participation' means that a forecaster filled out the portion of the survey that asked for his point forecast of inflation. Because we use the probabilistic forecasts in our analysis, hereafter 'participation' means that a forecaster filled out the portion of the survey that asked for his probabilistic forecast.

[3] See http://www.philadelphiafed.org/files/spf/survq107.html.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

While it is easy to construct numerical examples of the above sort, careful empirical analysis is required to evaluate how forecaster heterogeneity and changing panel composition may affect aggregated predictions in actual surveys. To shed light on the matter, we have examined the data on probabilistic inflation expectations obtained by the Survey of Professional Forecasters (SPF) in the period 1992–2006. After describing the SPF data in Section 2, this paper presents our findings in Sections 3 and 4.

We conclude that the interpretative problem is always serious in principle and is often serious in practice. Three factors contribute to this conclusion. First, we show in Section 3 that the predictions reported by SPF panel members exhibit considerable heterogeneity. Moreover, this heterogeneity exhibits strong persistence. That is, forecasters who are relatively uncertain about future inflation in one survey tend to be relatively uncertain throughout their participation in the panel. Those who expect high inflation in one survey tend to expect high inflation in other surveys. Thus it appears that the heterogeneity observed in the SPF forecasts arises out of permanent differences between forecasters in the way that they form inflation expectations.

Second, we show in Section 4 that the composition of the panel changes substantially over time, in part due to long-run turnover in the forecasters who officially serve as members of the panel and in part due to short-run variation in the panel members who actually respond to the survey. Consider, for example, the year-to-year stability of the panel. On average, 34 forecasters participated in each quarterly administration of the SPF during the 1992–2006 period. However, an average of nine forecasters who participated in a given quarter did not participate four quarters later, with another 10 or so taking their place at that time. Thus, when comparing predictions made four quarters apart, one confronts the problem that an average of 43 or 44 forecasters participated in at least one of the two surveys, but only about 25 participated in both.

Third, we report in Section 4 that little is known about the process that determines panel composition. Time series analysis of aggregated predictions would be a well-defined inferential problem if it were credible to assume that panel members are randomly recruited from a stable population of potential forecasters and that participation in the survey after recruitment is statistically independent of forecasters' beliefs about inflation. Then data on forecaster expectations would be missing completely at random. However, evidence to justify these assumptions is not available.

One can argue that the assumption that data are missing completely at random greatly simplifies analysis and hence should be maintained unless there exists persuasive evidence that this assumption is incorrect. In the absence of knowledge of the SPF recruitment and participation process, this assumption is not refutable. Thus a researcher who wishes to assume that SPF data are missing completely at random cannot be proved wrong.

Whereas this argument hopes for the best, we work from a more conservative perspective. As we see it, the underlying difficulty is that the changing composition of the SPF panel creates a problem of partial identification due to missing data (see Manski, 2007). Without knowledge of the forecaster participation process, one can only bound the distribution of inflation expectations in the panel at a given point in time, and similarly bound the distribution of changes in expectations over time. The constantly changing composition of the SPF panel implies that a large fraction of the relevant data is typically missing. Hence the bounds are quite wide.

In the absence of knowledge of the process determining panel composition, we recommend against the traditional use of the time series of aggregated SPF predictions to measure the evolution of forecasters' expectations. While other authors have advised against using the consensus forecasts, to our knowledge we are the first to argue that changing panel composition should discourage researchers from using the time series of consensus forecasts. Such time series conflate

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

changes in the expectations of individual forecasters with changes in the composition of the SPF panel.[4] Disentangling the two factors requires knowledge of the forecaster participation process.

Keane and Runkle (1990) also argue against using consensus forecasts. First, interpreting point forecasts as conditional expectations, they point out that the average of several expectations conditioning on different information sets need not be the conditional expectation given any one information set. Second, they argue that consensus forecasts 'mask' forecaster heterogeneity. They conclude: 'for both of these reasons ... researchers must use *individual* data in order to test hypotheses about how people form expectations.'

To replace analysis of aggregated predictions, we too recommend study of the time series of the predictions made by individual forecasters. As a prelude, we introduce in Section 3 a straightforward and appropriate way to describe the cross-sectional heterogeneity of predictions in a given survey. We consider each forecaster separately and compute parameters that measure the central tendency and spread of the elicited subjective probability distribution for future inflation; in particular, we suggest the subjective median and interquartile range.[5] This done, a plot showing the subjective (median, IQR) of each forecaster clearly portrays the heterogeneity of inflation forecasts at a point in time.[6]

To describe the evolution of expectations across the quarterly administrations of the survey, in Section 4 we recommend enhancing the plot with arrows to indicate how each forecaster changes his beliefs from one quarter to the next. Although we think that study of the predictions made by individual forecasters is the most appropriate way to use the SPF data, some researchers may continue to seek the relative simplicity of a single time series of aggregated predictions. For these researchers, Section 4 shows how to perform modest forms of time series analysis with sub-panels of fixed composition.

Although the data analysis in this paper focuses on probabilistic inflation expectations in the SPF, we emphasize in the concluding Section 5 that the themes developed here apply much more broadly. They apply equally well to the other probabilistic and point forecasts obtained in the SPF and, moreover, to the other panels of macroeconomic forecasters that we listed at the outset. Indeed, they apply even more broadly to panels making other types of forecasts. A prominent example outside of economics is the Intergovernmental Panel on Climate Change (IPCC).[7]

Before proceeding, we call the reader's attention to an alternative and very different approach to coping with entry and exit of forecasters, proposed by Capistran and Timmermann (2009) in a paper written contemporaneously and independently of our own. They seek combinations of

[4] In 2004 the SPF staff published a memo acknowledging this problem (see https://www.phil.frb.org/files/spf/WebMemo.pdf). In the memo, the staff considered creating an 'experimental panel' of forecasters who had participated in the SPF continuously since 1999. However, they found that no forecaster participated in every survey between 1Q1999 and 2Q2004. Hence the contemplated experimental panel would have had no members.

[5] The SPF elicits subjective probabilities that the inflation rate will fall into each of 10 intervals, rather than full subjective distributions. Hence an auxiliary assumption about the distribution of probability mass within each interval is needed to compute exact values for these parameters. The SPF intervals are narrow, so this is not much of a concern in practice. See Section 2 for further discussion.

[6] We earlier suggested this idea in our working paper (Engelberg *et al*., 2006, Section 5) and develop it more fully here. Engelberg *et al*. (2009) is a revised version of Sections 1–4 of the 2006 working paper, but does not include the material in Section 5. The present paper builds on and supersedes Section 5.

[7] The IPCC was established by the World Meteorological Organization and the United Nations Environment Programme in 1988. Its role is 'to assess on a comprehensive, objective, open and transparent basis the scientific, technical and socioeconomic information relevant to understanding the scientific basis of risk of human-induced climate change, its potential impacts and options for adaptation and mitigation' (http://www.ipcc.ch/pdf/ipcc-principles/ipcc-principles.pdf). The panel surveys scientific articles on climate change and summarizes the findings in 'assessment reports' that are released every few years. Among other things, the reports contain the panel members' aggregated beliefs that increases in global temperatures are due to human activity. The composition of the panel of climatologists writing the assessment reports can impact the beliefs that are reported in the reports. Comparison of the beliefs expressed in different reports may therefore be problematic: temporal variation in aggregated beliefs could reflect changes in the composition of the authors of the reports rather than changes in the available information about climate change.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

available past forecasts that best predict subsequent realizations of the quantity forecast, given specified prediction criteria. They do not pose assumptions on the entry–exit process that provide a foundation for use of the estimated best predictors to forecast future realizations. It may be that assumption of some form of temporal stability in the entry–exit process would provide such a foundation.

2. THE SPF DATA

The Survey of Professional Forecasters has been administered since 1990 by the Federal Reserve Bank of Philadelphia. The SPF was begun in 1968 by the American Statistical Association and the National Bureau of Economic Research; hence it was originally called the ASA-NBER survey. The panel of forecasters, who include university professors and private sector macroeconomic researchers, are asked to predict American GDP, inflation, unemployment, interest rates, and other macroeconomic variables.[8] The survey, which is performed quarterly, is mailed to panel members the day after government release of quarterly data on the national income and product accounts.

2.1. Question Format

Each quarter, the SPF asks panel members to make point and probabilistic forecasts of annual real GDP and inflation. To analyze the responses, it is important to understand the specific format of the questions. We describe here the format of the probabilistic forecasts for inflation, which are the focus of this paper.

In the four quarterly surveys administered during calendar year t, respondents are asked to forecast the percentage change in the GDP price index between the ends of years $t-1$ and t. They are also asked to forecast the corresponding change in the price index between the ends of years t and $t+1$. In each case, the SPF instrument partitions the real number line into intervals and asks respondents to report their subjective probabilities that inflation will take a value in each interval. During the sample period that we study, the intervals are $(-\infty, 0)$, $[x, x+1)$ for $x = 0, 1, \ldots, 7$ and $[8, \infty)\%$.

As described above, forecasts are made quarterly for the current year and the next year. To identify which year is being forecast and when the forecast is made, we will write that a forecast is 'X quarters ahead for year Y'. For example, in the first quarter of 1995 forecasts were made for inflation in 1995 and 1996. We label these two forecasts respectively as a four-quarter-ahead forecast for 1995 and an eight-quarter-ahead forecast for 1996. Thus, within year t, we observe eight forecasts made at different horizons: one quarter ahead for year t, ..., four quarters ahead for year t, five quarters ahead for year $t+1$, ..., and eight quarters ahead for year $t+1$.

2.2. Sample for Analysis

The SPF began in 1968, but we restrict attention to data collected from 1992 on. There are two main reasons for this:

1. Our analysis will focus on individual forecasters, following their forecasts over time. Before the Philadelphia Fed began to administer the survey, there is some evidence that forecaster IDs

[8] A partial list of respondents is posted in the Philadelphia Fed's quarterly release at http://www.phil.frb.org/econ/spf/.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

were reused over time as some forecasters left the panel and others joined. The Fed took over the survey in Quarter 3 of 1990, and has ensured since that forecaster IDs are not reused.[9]

2. The intervals on which respondents place probabilities have changed over the years. There were six intervals from Quarter 1 of 1983 through the end of 1991. There have been 10 intervals since then.

Because of these issues, we restrict our analysis to the survey responses from Quarter 1 of 1992 through Quarter 4 of 2006. As Table I demonstrates, even after this restriction our sample is large, with 4038 observations provided by 140 unique forecasters over the 15-year period.

2.3. Using the Probabilistic Forecasts to Estimate Subjective Distributions

Each SPF probabilistic forecast elicits the subjective probability that inflation will fall in 10 intervals. Hence the forecast does not fully reveal a respondent's subjective distribution. Suppose, for example, that a forecaster reports a 0.3 probability that inflation will be in the interval [2, 3)%, a 0.6 probability for the interval [3, 4) and a 0.1 probability for the interval [4, 5). Then we can infer these points on the forecaster's cumulative distribution function: $F(0.02) = 0$, $F(0.03) = 0.3$, $F(0.04) = 0.9$, and $F(0.05) = 1$.

To compute precise values for parameters measuring the central tendency and spread of each subjective distribution, we need to assume how probability mass is distributed within each interval. We assume that the mass is distributed uniformly within each interior interval. This assumption is relatively innocuous because the widths of the interior intervals are only 1% each.

We assume that probability mass placed in the tail intervals $(-\infty, 0)$ and $(8, \infty)$ is uniformly distributed within the intervals $(-1, 0)$ and $(8, 9)$ respectively. This assumption could in principle be consequential, but it is innocuous in our work for three reasons. First, relatively few respondents place any mass in the tail intervals. Second, those respondents who do use the tail intervals generally place only small mass in them. Third, our analysis uses the median of a subjective distribution to measure central tendency and the interquartile range (IQR) to measure spread. These parameters are robust to variation in the distribution of probability mass within the tail intervals.

Table I. Descriptive statistics

Quarters ahead	Observations	Missing data	Unique forecasters	Mean forecasters per survey
1	507	63	117	33.8
2	513	52	124	34.2
3	527	51	129	35.1
4	482	56	109	32.1
5	508	62	116	33.9
6	507	58	124	33.8
7	522	56	126	34.8
8	472	66	108	31.5
All	4038	464	140	33.7

Note: We count an observation as missing if the forecaster does not provide values for his subjective distribution.

[9] The Philadelphia Fed still must decide whether a forecaster ID should follow a forecaster when he changes employer. Information on the Fed's web site indicates that such decisions are based on judgments as to whether the forecasts represent the firm's or the individual's beliefs. See http://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/spf-documentation.pdf.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

An alternative analytical approach would be to make no assumptions about the distribution of probability mass within each interval. In that case, one cannot compute precise values for the median and IQR of a subjective distribution but one can obtain bounds on these parameters. This nonparametric approach was used in Engelberg *et al.* (2009), in a study of the central tendency of forecasters' expectations.

3. CROSS-SECTIONAL HETEROGENEITY IN INFLATION EXPECTATIONS

The median of a forecaster's subjective probability distribution measures the central tendency of his beliefs, and the interquartile range measures the uncertainty that this forecaster perceives. A simple way to summarize the cross-sectional distribution of forecaster beliefs at a point in time is to create a two-dimensional plot with subjective median on one axis and IQR on the other. To illustrate, the 10 plots in Figure 1 show the four-quarter-ahead inflation forecasts made in the first quarter of each year from 1997 through 2006. The vertical line in each plot locates the cross-sectional median of forecasters' subjective medians. The horizontal line locates the cross-sectional median of their subjective IQRs.

The plots are simple to interpret. Each point in a plot represents a unique forecaster. The intersection of the lines in a plot give the expectations of a hypothetical 'median forecaster'. When the points cluster towards the top, forecasters tend to feel much uncertainty. When the points are dispersed horizontally, disagreement in the central tendency of forecasts is high.

For example, in the first quarter of 2004, the subjective median of almost all forecasters for the inflation rate in the year ahead lay in a tight band between 1% and 2%, with a cross-sectional median of about 1.5%. Subjective IQRs ranged between 0.5% and 1.5%, with a cross-sectional median of about 1.2%.[10] Thus there was remarkably little disagreement in the central tendency of the SPF inflation forecasts. Forecasters varied moderately in their uncertainty about inflation in the year ahead.

Two years later, in the first quarter of 2006, the plot looks strikingly different. At this point in time, the subjective median varied considerably across forecasters, from a low of about 1.8% to a high of about 4.4%, with a cross-sectional median of 2.6%. Subjective IQRs ranged between 0.5% and 2.5%, with a cross-sectional median of 1%. Thus there was much more heterogeneity in inflation expectations in 2006 than in 2004.

The plots shown in Figure 1 portray the SPF data very differently from the quarterly summaries of findings released by the Philadelphia Federal Reserve Bank. These releases aggregate forecasters' probabilistic predictions by reporting the cross-sectional means of their elicited subjective probability distributions. As recognized by Giordani and Soderlind (2003), the cross-sectional mean of these distributions is a hybrid statistic that conflates forecaster uncertainty and disagreement. To illustrate, suppose that there are two forecasters, labeled A and B, and consider two scenarios. In one scenario, forecaster A places all probability mass in the inflation interval [2, 3)%, while forecaster B places all probability mass in the inflation interval [3, 4)%. In the other scenario, both forecasters place half their probability mass in the interval [2, 3) and half in the interval [3, 4). In the first scenario, the forecasters are individually quite certain about the outcome but they completely disagree with one another. In the second scenario, the forecasters are individually uncertain about the outcome but they completely agree. Reporting the mean probability mass in each interval makes these two scenarios indistinguishable.

We recommend that the Philadelphia Fed include plots like Figure 1 in the quarterly summaries of SPF findings reported to the public. The plots show, in a transparent manner, an informationally

[10] The smallest possible IQR is 0.5%. This occurs when a forecaster places all of his probability mass in a single interval. Our assumption that mass is distributed uniformly within each interval then implies that the IQR is 0.005.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

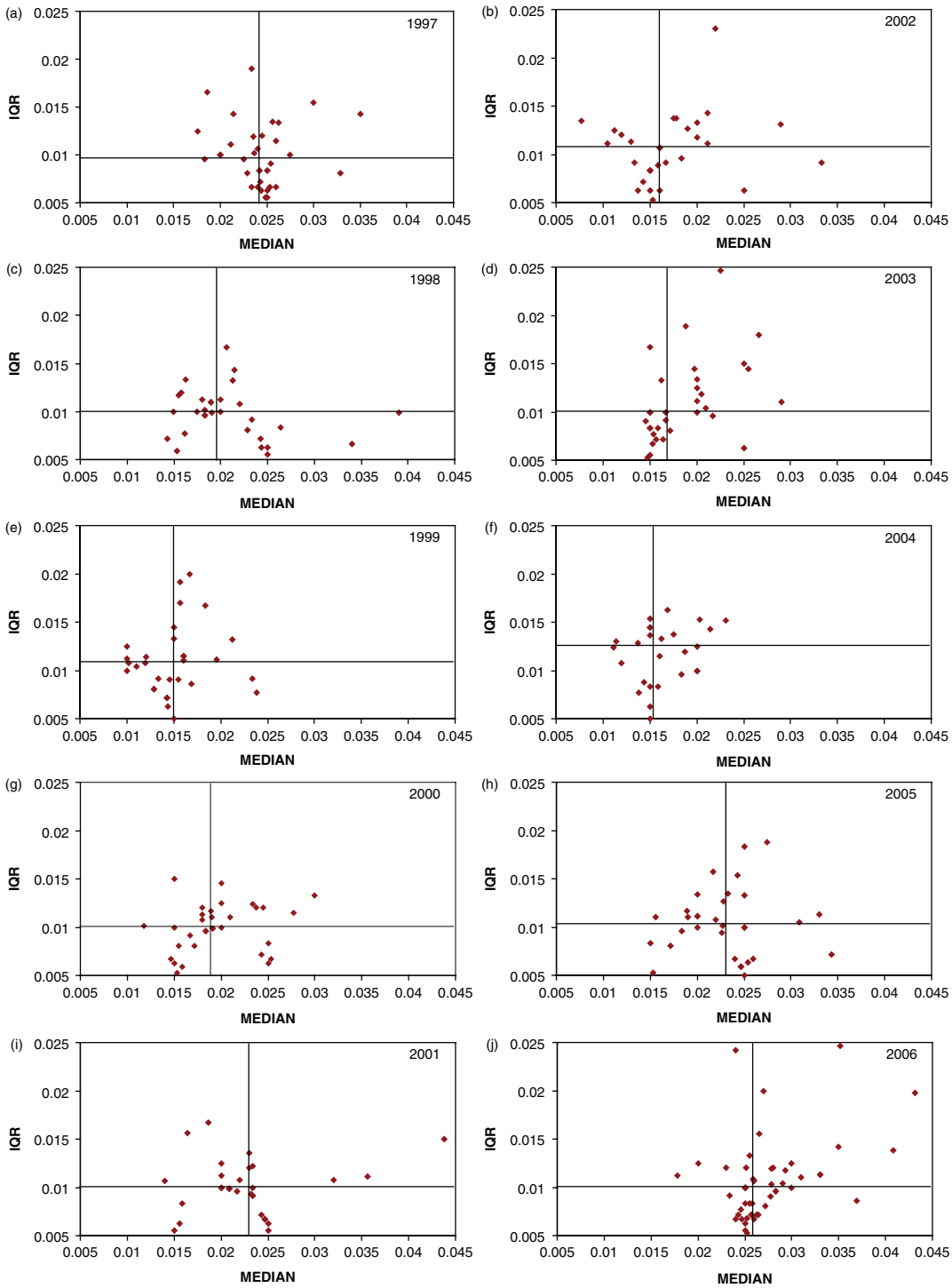

Figure 1. Median/IQR plot for four-quarter-ahead inflation forecasts 1997–2006. This figure is available in color online at wileyonlinelibrary.com/journal/jae

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

rich summary of the predictions made by the panel of forecasters. Scanning across the x-axis, one observes the degree to which forecasters agree or disagree with one another in the central tendencies of their forecasts. Scanning across the y-axis, one observes the uncertainty that forecasters perceive about future inflation. Scanning both axes jointly, one observes the association between central tendency and spread in the individual forecasts.

3.1. Persistence of Heterogeneity

For time series analysis of the SPF data, it is important to know whether the cross-sectional heterogeneity apparent in Figure 1 is a transient phenomenon or persists across surveys. In particular, we want to know whether panel members whose forecasts have high (low) subjective median or IQR in one survey tend to have high (low) median or IQR in other surveys.

Table II considers all inflation forecasts observed during our sample period and displays the serial and contemporaneous correlation of forecasters' subjective medians and IQRs. Table II presents both arithmetic and rank correlations. The arithmetic correlations are computed on the raw data. The rank correlations are computed by transforming each raw forecast into its rank within the group of contemporaneous forecasts made by panel members, and then computing arithmetic correlations on the transformed data.

We find strikingly strong long-term persistence in the IQR values. The arithmetic (rank) correlation between forecasters' subjective IQR in surveys 1 year apart is 0.59 (0.50) and the arithmetic (rank) correlation across surveys 4 years apart is almost as large, being 0.55 (0.44). Thus some forecasters are persistently confident in their inflation forecasts and others are persistently cautious.[11]

We also find persistence in forecasters' subjective medians, although not as large in magnitude. The arithmetic (rank) correlation across surveys 1 year apart is 0.48 (0.32) and the arithmetic (rank) correlation across surveys 4 years apart is 0.11 (0.13). Thus some forecasters persistently expect higher inflation than others.

Table II also shows the correlations between forecasters' subjective medians and IQRs. We find a moderate positive arithmetic (rank) correlation of 0.18 (0.14) between contemporaneous medians

Table II. Correlations of subjective medians and IQRs

	Panel A: arithmetic correlation		Panel B: rank correlation	
	This quarter IQR	This quarter median	This quarter IQR	This quarter median
This quarter IQR	1	0.18	1	0.14
This quarter median	0.18	1	0.14	1
Last quarter IQR	0.55	0.1	0.59	0.11
Last quarter median	0.12	0.69	0.1	0.48
1 year ago IQR	0.59	0.11	0.5	0.11
1 year ago median	0.14	0.48	0.1	0.32
2 years ago IQR	0.56	0.1	0.47	0.08
2 years ago median	0.09	0.18	0.09	0.23
3 years ago IQR	0.53	0.13	0.42	0.1
3 years ago median	0.02	0.09	0.05	0.18
4 years ago IQR	0.55	0.19	0.44	0.12
4 years ago median	0.01	0.11	0.07	0.13

[11] In computations not shown in Table II, we have also found that the subjective IQR of inflation forecasts is strongly and persistently associated with the subjective IQR of forecasts of growth in GDP. Thus forecasters who are uncertain about future inflation tend to also be uncertain about future GDP growth.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

and IQRs, and this persists when one variable or the other is lagged. Thus forecasters who expect higher inflation tend to be more uncertain.

Overall, Table II suggests that the cross-sectional heterogeneity evident in the plots of Figure 1 arises out of permanent differences between forecasters in the way that they form inflation expectations. We think that it would be of great interest in future research to dig deeper and try to infer the distinct processes of expectations formation that different forecasters use.

4. MEASURING TIME SERIES VARIATION IN INFLATION EXPECTATIONS

4.1. Tracking Individual Forecasters

A direct and transparent way to study time series variation in the inflation expectations of the SPF panel is to track the responses of individual forecasters across the quarterly surveys in which they participate. Considering each forecaster in isolation shows clearly how individual expectations evolve over time. Comparison of the time paths of the responses of different forecasters illuminates the heterogeneity of the SPF panel.

Figure 2 gives two illustrations. The top figure displays the subjective medians and IQRs for 2001 GDP growth elicited from forecasters who participated in the SPF in both 3Q2001 and 4Q2001. Thus the figure shows GDP growth expectations before and after the terrorist attacks of 11 September 2001. Each forecaster is depicted by an arrow whose tail is his 3Q2001 prediction and whose tip is his 4Q2001 prediction. The figure shows that nearly all forecasters revised their

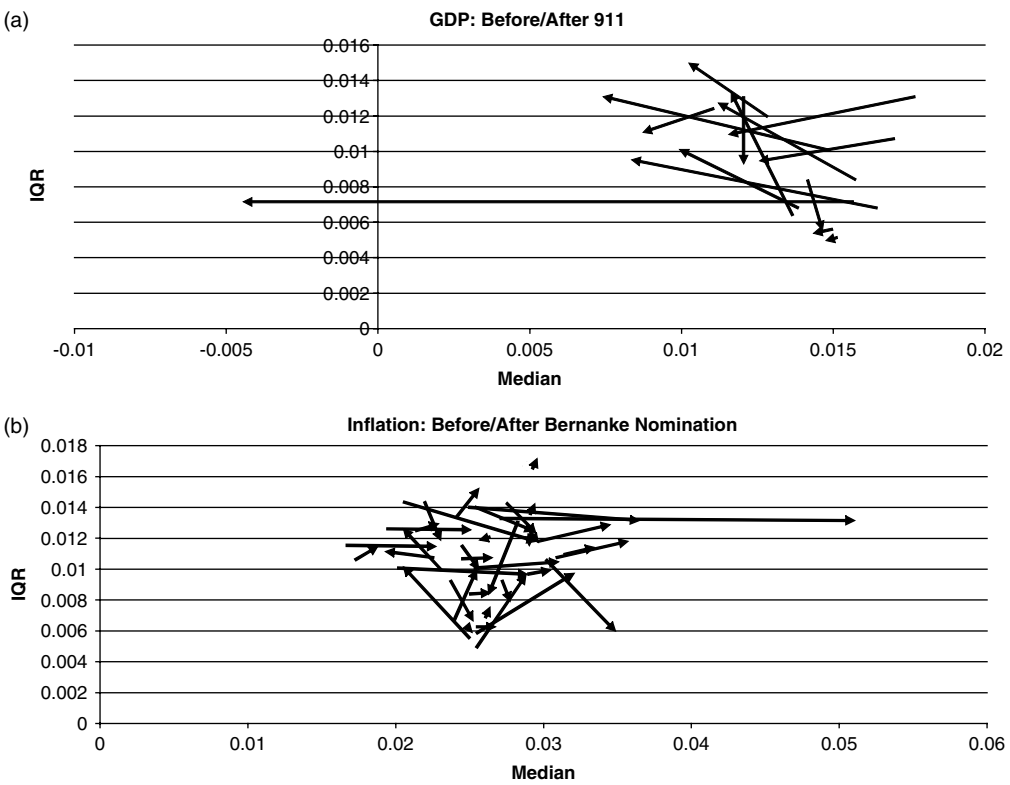

Figure 2. Change in expectations around events

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

subjective medians downward between 3Q2001 and 4Q2001. However, forecasters varied in the direction of revisions to their subjective IQRs, with some becoming more certain about output growth and others becoming less certain.

The bottom figure displays the subjective medians and IQRs for 2006 inflation elicited from forecasters who participated in the SPF in both 3Q2005 and 4Q2005. Thus the figure shows inflation expectations before and after Ben Bernanke's nomination to be Chair of the Board of Governors of the Federal Reserve System on 24 October 2005. The figure shows that most forecasters revised their subjective medians upwards but they varied considerably in the magnitude of the revision. Revisions to subjective IQRs were very heterogeneous in both direction and magnitude.

We recommend that the Philadelphia Fed include plots like Figure 2 in the quarterly summaries of SPF findings reported to the public. The plots show, in a transparent manner, an informationally rich summary of the quarter-to-quarter revisions to the predictions made by the panel of forecasters.

While Figure 2 nicely displays revisions to SPF expectations from quarter to quarter, it is not feasible to show longer time series in the same manner. Figure 3 shows the complete time series of forecasts made by two panel members from 2000 through 2006. Comparing the two panels, we observe that forecaster #472 is persistently much more certain about future inflation than is forecaster #483. The former panel member always places all of his probability mass in one, two or three intervals, implying IQRs that range from 0.5% to 1%. In contrast, the latter person never uses fewer than three intervals and often places mass in all ten; his IQRs range from 0.63% to 2.9%. We also observe that forecaster #472 persistently forecasts lower future inflation than does #483. In almost every period when both forecasters complete the survey, the subjective median for forecaster #472 is less than that for #483.

The two forecasters described in Figure 3 clearly differ in their beliefs about the ability of the Federal Reserve System to keep inflation within the Fed's 'comfort zone'. The Fed has not announced an explicit target range for inflation, as have the central banks of some other countries. However, many observers think that, since the turn of the century, the Fed has had an implicit target of 1–3% inflation, sometimes called the comfort zone.[12] With this background, it is of interest to learn how probable forecasters think it is that inflation will remain in this range.

Consider forecaster #472. Throughout most of the period 2000–2006, this person concentrated all of his probability mass for future inflation in the range 1–3%. There were only two minor exceptions. In the period 2000–2003, he sometimes placed small probability on the event that future inflation would be in the range 0–1%. During part of the year 2006, he placed high probability on the event that inflation in 2006 would be in the range 3–4%. However, even in this period, he believed that any rise in the inflation rate would be temporary. His five-quarter-ahead to eight-quarter-ahead forecasts in 2006 were that inflation in 2007 would certainly be in the range 1–3%.

Forecaster #483 plainly is much less sure that inflation will remain in the range 1–3%. Of the 43 forecasts that he made in the period 2000–2006, in only 10 cases did he place more than 60% chance on inflation being in this range. He regularly placed moderate probability on the event that the inflation rate would be above 4%.

4.2. Analysis of Aggregated Predictions: The Problem of Temporal Variation in Panel Composition

The only downside to tracking individual forecasters as in Figure 3 is that the data are too rich to assimilate with ease. As shown earlier in Table I, a total of 140 forecasters were members of the panel during some part of the period 1992–2006, with an average of 33.7 participating per

[12] We are grateful to Matthew Shapiro for his insights on this subject.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Year	Quarters Ahead Forecast	FORECASTER ID # 472 Probability Mass Placed in Each of 10 Bins from Lowest (left) to Highest (right)										Median	IQR
2000	8		10%	70%	20%							1.57%	0.71%
2000	7		0%	90%	10%							1.56%	0.56%
2000	6		5%	80%	15%							1.56%	0.63%
2000	5		20%	65%	15%							1.46%	0.77%
2000	4			55%	45%							1.91%	0.99%
2000	3			30%	70%							2.29%	0.81%
2000	2			0%	100%							2.50%	0.50%
2000	1	Did Not Participate / Left Blank											
2001	8			55%	45%							1.91%	0.99%
2001	7		5%	70%	25%							1.64%	0.71%
2001	6			40%	60%							2.17%	0.96%
2001	5	Did Not Participate / Left Blank											
2001	4		5%	80%	15%							1.56%	0.63%
2001	3			35%	65%							2.23%	0.90%
2001	2	Did Not Participate / Left Blank											
2001	1			10%	90%							2.44%	0.56%
2002	8			70%	30%							1.71%	0.81%
2002	7			50%	50%							2.00%	1.00%
2002	6	Did Not Participate / Left Blank											
2002	5		5%	80%	15%							1.56%	0.63%
2002	4	Did Not Participate / Left Blank											
2002	3		20%	80%								1.38%	0.63%
2002	2	Did Not Participate / Left Blank											
2002	1		5%	95%								1.47%	0.53%
2003	8	Did Not Participate / Left Blank											
2003	7		10%	70%	20%							1.57%	0.71%
2003	6	Did Not Participate / Left Blank											
2003	5		15%	80%	5%							1.44%	0.63%
2003	4	Did Not Participate / Left Blank											
2003	3	Did Not Participate / Left Blank											
2003	2	Did Not Participate / Left Blank											
2003	1			100%								1.50%	0.50%
2004	8	Did Not Participate / Left Blank											
2004	7	Did Not Participate / Left Blank											
2004	6	Did Not Participate / Left Blank											
2004	5			90%	10%							1.56%	0.56%
2004	4	Did Not Participate / Left Blank											
2004	3			10%	90%							2.44%	0.56%
2004	2	Did Not Participate / Left Blank											
2004	1			10%	90%							2.44%	0.56%
2005	8	Did Not Participate / Left Blank											
2005	7			25%	75%							2.33%	0.67%
2005	6	Did Not Participate / Left Blank											
2005	5			75%	25%							1.67%	0.67%
2005	4			95%	5%							1.53%	0.53%
2005	3			30%	70%							2.29%	0.81%
2005	2			1%	98%	1%						2.50%	0.51%
2005	1				100%							2.50%	0.50%
2006	8			95%	5%							1.53%	0.53%
2006	7			45%	55%							2.09%	0.99%
2006	6			10%	90%							2.44%	0.56%
2006	5			50%	50%							2.00%	1.00%
2006	4				95%	5%						2.53%	0.53%
2006	3				5%	95%						3.47%	0.53%
2006	2				15%	85%						3.41%	0.59%
2006	1				70%	30%						2.71%	0.81%

Year	Quarters Ahead Forecast	FORECASTER ID # 483 Probability Mass Placed in Each of 10 Bins from Lowest (left) to Highest (right)										Median	IQR
2000	8	5%	25%	35%	20%	10%	3%	2%				1.57%	1.70%
2000	7	Did Not Participate / Left Blank											
2000	6	Did Not Participate / Left Blank											
2000	5		15%	60%	15%	5%	3%	2%				1.58%	0.83%
2000	4	Did Not Participate / Left Blank											
2000	3		3%	15%	40%	25%	7%	5%	3%	2%		2.80%	1.51%
2000	2	Did Not Participate / Left Blank											
2000	1			35%	50%	15%						2.30%	1.09%
2001	8	Did Not Participate / Left Blank											
2001	7		5%	10%	25%	20%	12%	10%	7%	6%	5%	3.50%	2.90%
2001	6	Did Not Participate / Left Blank											
2001	5		5%	10%	25%	20%	12%	10%	7%	6%	5%	3.50%	2.90%
2001	4		5%	10%	25%	20%	12%	10%	7%	6%	5%	3.50%	2.90%
2001	3	Did Not Participate / Left Blank											
2001	2		5%	10%	35%	24%	20%	5%	1%			3.00%	1.76%
2001	1	Did Not Participate / Left Blank											
2002	8		5%	10%	25%	20%	12%	10%	7%	6%	5%	3.50%	2.90%
2002	7	Did Not Participate / Left Blank											
2002	6		7%	20%	30%	15%	8%	7%	6%	5%	2%	2.77%	2.48%
2002	5	Did Not Participate / Left Blank											
2002	4	1%	15%	30%	20%	15%	10%	5%	2%	1%	1%	2.20%	2.30%
2002	3	1%	15%	30%	20%	15%	10%	5%	2%	1%	1%	2.20%	2.30%
2002	2	1%	18%	50%	15%	12%	4%					1.62%	1.28%
2002	1	Did Not Participate / Left Blank											
2003	8	1%	7%	20%	30%	14%	8%	7%	6%	5%	2%	2.73%	2.53%
2003	7	1%	7%	28%	22%	14%	8%	7%	6%	5%	2%	2.64%	2.77%
2003	6	1%	15%	34%	18%	10%	8%	6%	4%	3%	1%	2.00%	2.44%
2003	5	Did Not Participate / Left Blank											
2003	4	1%	18%	35%	20%	17%	5%	4%				1.89%	1.89%
2003	3	1%	16%	40%	25%	15%	2%	1%				1.83%	1.52%
2003	2	1%	16%	40%	25%	15%	2%	1%				1.83%	1.52%
2003	1	Did Not Participate / Left Blank											
2004	8	1%	15%	34%	18%	10%	8%	6%	4%	3%	1%	2.00%	2.44%
2004	7	4%	15%	34%	18%	10%	8%	5%	3%	2%	1%	1.91%	2.22%
2004	6	4%	15%	34%	18%	10%	8%	5%	3%	2%	1%	1.91%	2.22%
2004	5	Did Not Participate / Left Blank											
2004	4		15%	60%	15%	10%						1.58%	0.83%
2004	3		15%	30%	40%	10%	5%					2.13%	1.42%
2004	2		10%	20%	40%	25%	5%					2.50%	1.45%
2004	1		5%	10%	70%	15%						2.50%	0.71%
2005	8	4%	15%	34%	18%	10%	8%	5%	3%	2%	1%	1.91%	2.22%
2005	7	1%	5%	20%	35%	20%	8%	5%	3%	2%	1%	2.69%	1.75%
2005	6	1%	1%	17%	35%	25%	10%	5%	3%	2%	1%	2.89%	1.67%
2005	5	1%	1%	17%	35%	25%	10%	5%	3%	2%	1%	2.89%	1.67%
2005	4		5%	15%	40%	15%	12%	8%	5%			2.75%	1.88%
2005	3		5%	16%	50%	15%	8%	5%	1%			2.58%	1.19%
2005	2		5%	15%	55%	15%	10%					2.55%	0.91%
2005	1			15%	80%	5%						2.44%	0.63%
2006	8	1%	1%	17%	35%	25%	10%	5%	3%	2%	1%	2.89%	1.67%
2006	7	1%	1%	17%	35%	25%	10%	5%	3%	2%	1%	2.89%	1.67%
2006	6	1%	1%	17%	35%	25%	10%	5%	3%	2%	1%	2.89%	1.67%
2006	5	1%	1%	17%	35%	25%	10%	5%	3%	2%	1%	2.89%	1.67%
2006	4		5%	12%	20%	25%	15%	12%	5%	3%	3%	3.52%	2.47%
2006	3		5%	12%	20%	25%	15%	12%	5%	3%	3%	3.52%	2.47%
2006	2		5%	12%	20%	29%	15%	8%	5%	3%	3%	3.45%	2.20%
2006	1		5%	12%	20%	25%	15%	12%	5%	3%	3%	3.52%	2.47%

Figure 3. Probabilistic forecasts of persistently certain and uncertain forecasters.

survey. Tracking so many separate time paths is burdensome. Hence it has been common to report the time series of statistics that aggregate forecasts across the SPF panel. Unfortunately, this is not straightforward to do.

Viewing Figure 1, one almost immediately wants to scan the 10 plots and draw conclusions about the time series variation in SPF inflation expectations from 1997 to 2006. A particularly simple summary of the time series is given by scanning the vertical and horizontal lines in the plots, which show how the subjective median and IQR of the 'median forecaster' evolve over the sample period.

The 'median forecaster' time series in Figure 1 would be interpretable if the SPF panel were a fixed group of forecasters who provide data in all surveys. However, the composition of the panel varies considerably over time. It varies from quarter to quarter because panel members often do not provide their requested forecasts. For example, the two forecasters considered in Figure 3 respectively did not provide 18 and 14 of the 56 forecasts requested in the period 2000–2006. The panel composition varies over the longer run when some forecasters leave the panel permanently and others are added.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

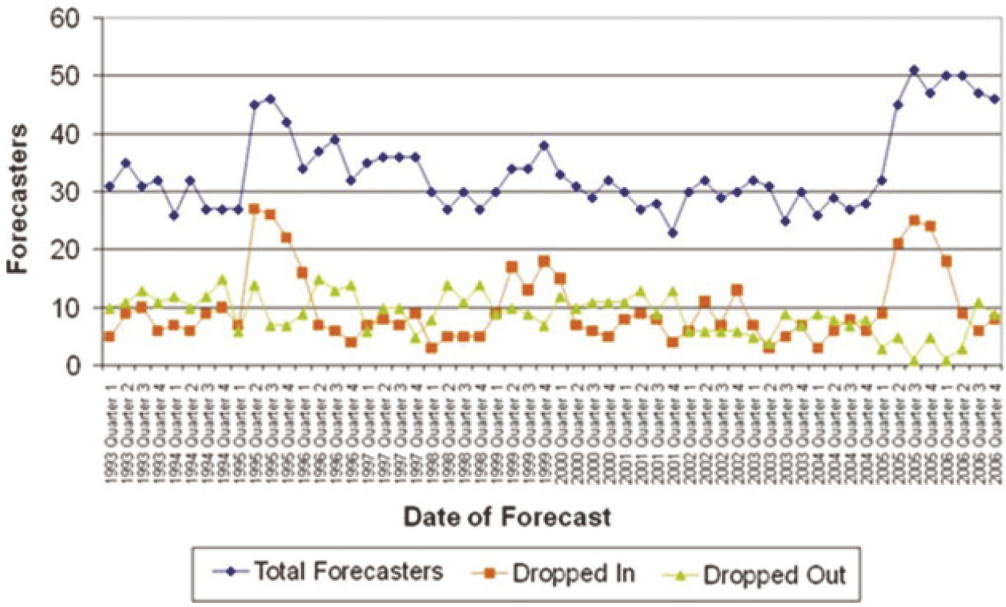

Figure 4. Forecaster participation in the SPF. The figure plots forecaster participation over our sample. 'Dropped In' plots the number of forecasters who participated in the current survey but not in the survey four quarters ago. 'Dropped Out' plots the number of forecasters who participated four quarters ago but did not participate in the current survey. This figure is available in color online at wileyonlinelibrary.com/journal/jae

Figure 4 shows the number of forecasters who participated in the SPF in each of the 56 quarterly surveys of our sample period, and the numbers of forecasters who 'dropped in' and 'dropped out' each quarter. We say that a forecaster has dropped in if he participates in the current survey but did not participate four quarters ago. A forecaster has dropped out if he does not participate in the current survey but participated four quarters ago. The data in the figure show that, on average, 9.8 forecasters drop into the survey each quarter and 8.9 forecasters drop out.[13] On average, 33.7 forecasters participate in the survey. Thus there is substantial change in panel composition across surveys.

Another perspective on the time series of panel composition is obtained by taking the individual forecaster as the unit of observation and computing transition probabilities for survey participation within the period when the forecaster is a panel member. For these computations, we define a forecaster's period of panel membership to begin with the first quarter in which he participates in the survey and end with the last quarter of participation. Let $y(t) = 1$ if a panel member participates in the survey at quarter t and $y(t) = 0$ otherwise. Aggregating across all SPF forecasters and all quarters in which they were panel members, we find that panel members who participate in one survey are much more likely to participate in later surveys. However, the transition probabilities are distant from zero and one in all cases. In particular, we find these transition probabilities for participation in the next quarter's survey and in the survey four quarters later, conditional on participation status at quarter t:

$$P[y(t+1) = 1 | y(t) = 1] = 0.83, \qquad P[y(t+4) = 1 | y(t) = 1] = 0.83,$$
$$P[y(t+1) = 1 | y(t) = 0] = 0.50, \qquad P[y(t+4) = 1 | y(t) = 0] = 0.46$$

[13] Average dropins exceed average dropouts because the size of the SPF panel jumped sharply in early 2005 and remained at this higher level throughout 2005 and 2006.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

The constantly changing composition of the SPF panel would not be problematic for aggregate time series analysis if it were credible to assume that (a) the Philadelphia Fed effectively draws new panel members at random from some population of potential forecasters and (b) forecasters who join the panel miss surveys and leave the panel at random, in the sense that inflation expectations are statistically independent of participation in the panel. Given these assumptions, the forecasts that are observed each survey are a random sample of the potential forecasts for that survey. Hence time series analysis of the SPF data is a well-defined problem in statistical inference.

Unfortunately, we are aware of no foundation for assumptions (a) and (b). The available documentation on the SPF does not explain how the Philadelphia Fed draws new panel members. We have discussed the matter with Fed staff, and have obtained the impression that the selection process somehow reconciles the subjective views of Fed staff on the suitability of persons for panel membership with the willingness of persons to serve on the panel. Each quarter the Fed sends the survey to the panel members and records the responses of those who send the survey back. Neither we nor the Fed are aware of why some forecasters do not respond, although those who administer the survey speculate that reasons include workload and vacation. When the panel size becomes small, the Fed actively recruits new members by sending out invitations to forecasters identified in professional directories and elsewhere. This presumably explains why spikes in participation occur in Figure 2 at 2Q1995 and 2Q2005.

In the absence of knowledge of the process that generates participation in the SPF, we recommend that researchers refrain from using the time series of the plots in Figure 1, or derived measures such as the expectations of the 'median forecaster', to draw conclusions about the evolution of forecasters' expectations. Such time series conflate changes in the expectations of individual forecasters with changes in the composition of the SPF panel. Knowledge of the forecaster participation process is necessary to disentangle the two factors.

The persistent heterogeneity of SPF forecasters documented in Section 3 makes changing panel composition particularly problematic for aggregate time series analysis of the SPF data. Over time, relatively confident forecasters may be replaced by relatively cautious ones, or vice versa. Forecasters with relatively high inflation expectations may be replaced by ones with relatively low expectations, or vice versa. Without knowledge of whether these changes in panel composition occur randomly or systematically, interpretation of the aggregate time series is not possible.

The underlying difficulty is that the changing composition of the SPF panel creates a problem of partial identification due to missing data. In the absence of knowledge of the forecaster participation process, it is only possible to bound the distribution of inflation expectations within the panel at a given point in time. Similarly, it is only possible to bound the distribution of changes in expectations over time. See Horowitz and Manski (1998) and Manski (2003, 2007) for analysis giving the specific form of the bounds.

Importantly, the bounds increase in width with the prevalence of missing data. The constantly changing composition of the SPF panel implies that a large fraction of the relevant data is typically missing. Hence the bounds are quite wide.

4.3. Aggregate Analysis of Sub-panels of Fixed Composition

Although we think that the most appropriate way to use the SPF data is to study the predictions of individual forecasters, we expect that there will remain demand for some form of aggregate analysis, if only to simplify the presentation of findings. If aggregate analysis is to be undertaken, we urge that it recognize the problem of changing panel composition.

Various possibilities open up if one finds it credible to assume that SPF panel members are drawn at random from a stable population of potential forecasters and that forecasters who join

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

the panel miss surveys and leave the panel at random. One may then apply available approaches for statistical analysis of panel data. For example, Keane and Runkle (1990), Davies and Lahiri (1995, 1999), and Rich and Tracy (2003) estimate regression models with forecaster fixed effects. Such fixed effects could be used to predict how forecasters who miss some surveys would have responded had they participated. We do not pursue this idea here because, as discussed earlier, we lack the requisite knowledge of the process that generates participation in the SPF.

In the absence of knowledge of the participation process, we recommend modest forms of analysis that focus attention on sub-panels of fixed composition. We limit attention here to inference on the change in aggregated predictions across two time periods. In this context, the simplest sub-panel of fixed composition is the group of forecasters who participate in both surveys. We earlier considered such a sub-panel in Figure 2, which displayed the expectations of individual forecasters before and after certain events. A larger sub-panel of fixed composition is the group of forecasters who participate in at least one of the two surveys. This group includes forecasters with missing data on one survey.

For simplicity, in this section we restrict attention to the four-quarter-ahead forecasts. Let $N_t(1, 1)$ denote the group of forecasters who respond to the SPF at times t and $t+1$, let $N_t(1, 0)$ be those who respond at time t but not $t+1$, and let $N_t(0, 1)$ denote those who respond at $t+1$ but not t. One fixed group of interest is $N_{tI} \equiv N_t(1, 1)$, the *intersection* (I) of the forecasters who participate at times t and $t+1$. Another is $N_{tU} \equiv N_t(1, 1) \cup N_t(1, 0) \cup N_t(0, 1)$, the *union* ($U$) of the forecasters who participate in both surveys. Let y_{ti} denote a prediction of interest, say the subjective median or IQR that forecaster i holds for future inflation at time t. We observe $(y_{ti}, y_{(t+1)i})$ for $i \in N(1, 1)$, y_{ti} for $i \in N(1, 0)$, and $y_{(t+1)i}$ for $i \in N(0,1)$.

A well-specified aggregate analysis considers some parameter of the cross-sectional distribution of y and asks how this parameter changes over time within a group of fixed composition. For concreteness, let the parameter of interest be the cross-sectional median of y; that is, the response of the 'median forecaster'. Then one well-specified object of interest is $\Delta_{tI} \equiv med(y_{(t+1)}|i \in N_{tI}) - med(y_t|i \in N_{tI})$ and another is $\Delta_{tU} \equiv med(y_{(t+1)}|i \in N_{tU}) - med(y_t|i \in N_{tU})$. In contrast, traditional analysis of aggregated predictions presents findings on the ill-specified *composite* (C) quantity $\Delta_{tC} \equiv med[y_{(t+1)}|i \in N_{tI} \cup N(0, 1)] - med[y_t|i \in N_{tI} \cup N(1, 0)]$, which compares two distinct groups of forecasters.

The SPF data directly reveal Δ_{tI} and Δ_{tC}. Figure 5 plots the values when t is the first quarter of the past year, $t+1$ is the first quarter of the current year, and y is the subjective median or IQR for inflation in the current year. The x-axis of the figure gives the years $t+1$ and the y-axis gives the corresponding values of Δ. The figure shows that the time series for Δ_{tI} and Δ_{tC} are almost identical when y is the subjective median for future inflation. However, they noticeably diverge in some years when y is the subjective IQR. In particular, the Δ_{tI} plot indicates that subjective uncertainty about future inflation decreased in the periods 1998–1999 and 2001–2002, while the Δ_{tC} plot indicates that it increased in these periods.

The SPF data do not directly reveal Δ_{tU}, whose value depends in part on missing values of y. However, we can bound Δ_{tU} if we find it credible to assume that the unobserved values of y must lie in some range. See Manski (2007, ch. 2) for extensive discussion of this type of analysis of missing data.

There are various possibilities for specification of the range. In the SPF context, we find it attractive to directly bound the temporal changes $y_{(t+1)i} - y_{ti}$ and then use this bound to derive a bound on unobserved values of y. This approach is attractive because we have in Section 3 documented strong persistence of heterogeneity of expectations across the SPF forecasters. Hence it is reasonable to think that the temporal changes $y_{(t+1)i} - y_{ti}$ tend not to be too large.

Let us assume therefore that $K_{tL} \leq y_{(t+1)i} - y_{ti} \leq K_{tU}$ for some lower and upper bounds K_{tL} and K_{tU}. In the case of persons in $N_t(1, 0)$, this implies that the unobserved $y_{(t+1)i}$ lies in the range

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

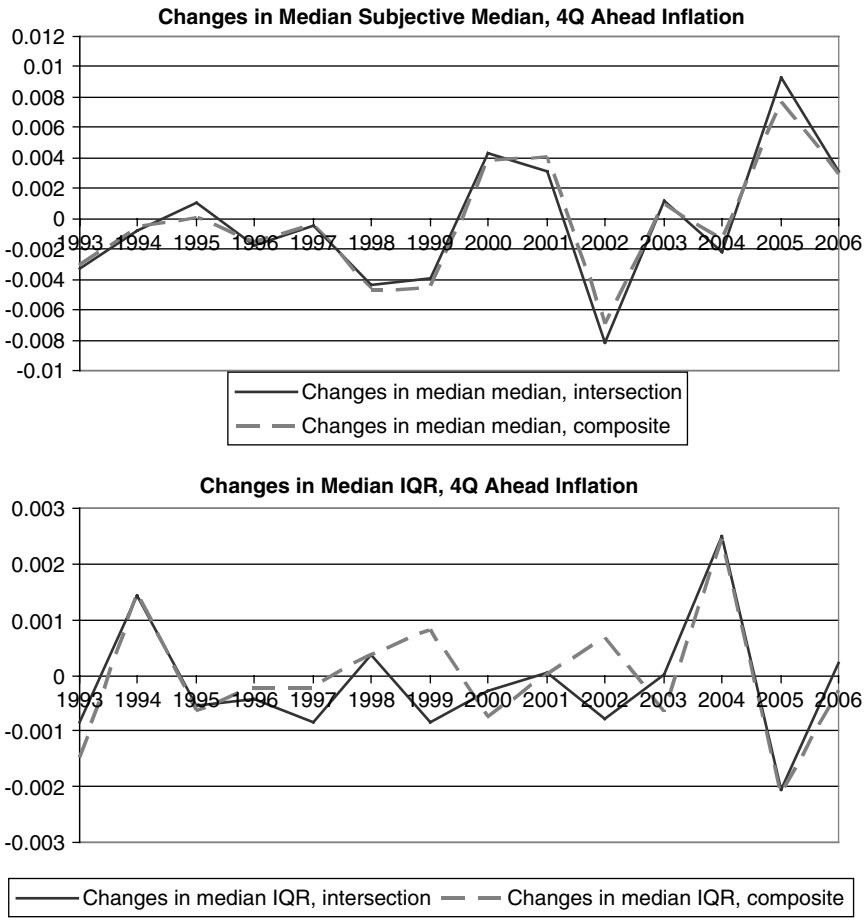

Figure 5. Compositional effects on changes in subjective medians and IQRs

$y_{ti} + K_{tL} \leq y_{(t+1)i} \leq y_{ti} + K_{tU}$. In the case of persons in $N_t(0, 1)$, this implies that the unobserved y_{ti} lies in the range $y_{(t+1)i} - K_{tU} \leq y_{ti} \leq y_{(t+1)i} - K_{tL}$. We think that a reasonable way to specify K_{tL} and K_{tU}, although certainly not the only sensible possibility, is to assume that unobserved changes in expectations are no larger than the largest observed changes. Thus we set

$$K_{tL} = [\min(y_{(t+1)i} - y_{ti}),\ i \in N_t(1, 1)]$$

and

$$K_{tU} = [\max(y_{(t+1)i} - y_{ti}),\ i \in N_t(1, 1)]$$

These assumptions yield lower and upper bounds on $y_{(t+1)i}$ for all members of $N_t(1, 0)$ and similarly yield bounds on y_{ti} for all members of $N_t(0, 1)$. These bounds in turn yield bounds on Δ_{tU}. The upper bound on Δ_{tU} is obtained by setting $y_{(t+1)i}$ at its upper bound for all members of $N_t(1, 0)$ and by setting y_{ti} at its lower bound for all members of $N_t(0, 1)$. The lower bound is similarly obtained by taking lower and upper bounds respectively.

Figure 6 plots Δ_{tC} and the bound on Δ_{tU}. Observe that the bound on Δ_{tU} uniformly covers the plot for Δ_{tC}. This is an empirical finding, not a logical necessity. In general, Δ_{tC} need not lie within our bound on Δ_{tU}.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

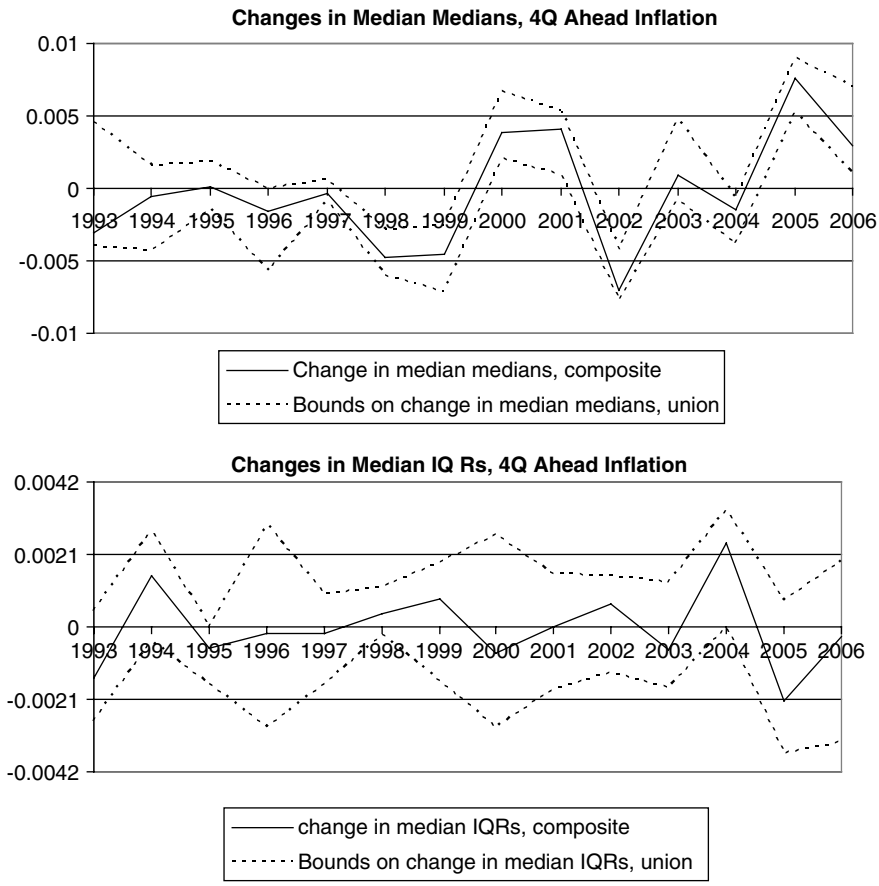

Figure 6. Bounds on changes in subjective medians and IQRs

In the top panels of Figures 5 and 6, we plotted Δ_{tL}, Δ_{tC}, and bounds for Δ_{tU}, where these measures were computed by letting y_{ti} equal the sample's subjective median. We chose to let y_{ti} represent forecaster i's subjective median (rather than his point forecast) because, unlike i's point forecasts, i's subjective median is well defined. See Engelberg *et al.* (2009) for an analysis of how SPF forecasters' point forecasts are related to their subjective medians and other measures of central tendency from the forecasters' subjective probability distributions.

Even though we prefer subjective medians to point forecasts, SPF forecasters' point forecasts still receive more attention than the measures of central tendency from their subjective probability distributions. We therefore produce new plots for Δ_{tL}, Δ_{tC}, and bounds for Δ_{tU} when y_{ti} is defined to be forecaster i's point forecast. See Appendix for these plots.

5. CONCLUSION

Surveys of professional forecasters are well-cited, important inputs for regulators and financial markets. Changes in expectations elicited via survey can have important policy implications and affect prices in asset markets. Given the influence such surveys have, it is important that we correctly infer changes in expectations from survey data.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

This paper has called attention to the problem of changing panel composition in surveys of forecasters, has documented the extent of the problem in the Survey of Professional Forecasters, and has recommended new ways to report and study the SPF data. To study the temporal variation of forecasts, we think that it is most informative to examine the time series of predictions made by individual forecasters, using plots like Figure 2 and presentations like Figure 3. This makes transparent the heterogeneity of the SPF panel and avoids improper inferences due to changing panel composition. Should analysis of aggregated predictions be thought desirable as a simplifying device, we recommend modest analyses of sub-panels of fixed composition. We warn against the traditional practice of aggregate time series analysis, which conflates changes in the expectations of individual forecasters with changes in the composition of the SPF panel.

Readers may have noticed that this paper reports no formal statistical analysis of the SPF data. This is because we lack the requisite knowledge of the sampling process recruiting forecasters into the panel and of the participation process among those who join. We could have performed statistical analyses under conventional assumptions of random sampling from a stable population of forecasters and random absence of forecasters following recruitment into the panel. However, we have found no foundation for these assumptions. Hence we think that conventional statistical analysis is meaningless in this context.

Although we have focused attention on the SPF, we reiterate that the themes developed here apply equally well to other panels of forecasters. In Section 4.2 we showed that, on average, 83% of the SPF forecasters who participate one quarter participate again four quarters later. This change in panel composition is not unique to the SPF. Through communications with conductors of the Duke/CFO Magazine Global Business Outlook Survey, the CESifo World Economic Survey, and the CBI SME (Small and Medium Enterprises), we have gathered that one of the surveys is unable to compute the turnover in panel composition due to respondent anonymity, and the other two surveys estimate that '85–90%' (for one survey) and 60% (for the other) of the respondents who participate in one round of the survey do so again in the next round of the survey.

Hence the methods we propose for dealing with the SPF data are relevant for other prominent surveys as well. Changing panel composition is a pervasive problem in surveys of professional forecasters, as is lack of knowledge of the participation process. The SPF is not exceptional in either respect.

Some readers of this paper may not find its conservative perspective congenial. We observed in the Introduction that, in the absence of knowledge of the forecaster recruitment and participation process, the assumption that data are missing completely at random is not refutable. Hence one might argue that this simplifying assumption should be maintained until evidence to the contrary emerges. To forestall endless debate about the validity of this or other simplifying assumptions, we see a strong need for research that sheds light on the forecaster recruitment and participation process. Only then will it become possible to reach consensus on the seriousness of the composition issue in survey response.

ACKNOWLEDGEMENTS

We thank Tom Stark of the Philadelphia Federal Reserve Bank for answering questions about the Survey of Professional Forecasters and John Graham for answering questions about the Duke/CFO survey. We thank Matthew Shapiro, two referees, and the editor for comments.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

REFERENCES

Baghestani H. 1994. Evaluating multiperiod survey forecasts of real net exports. *Economic Letters* **44**: 267–72.

Baghestani H. 2006. Federal Reserve vs. private forecasts of real net exports. *Economics Letters* **91**: 349–353.

Ball L, Croushore D. 2003. Expectations and the effects of monetary policy. *Journal of Money, Credit and Banking* **35**: 473–484.

Campbell SD. 2007. Macroeconomic volatility, predictability, and uncertainty in the Great Moderation: evidence from the Survey of Professional Forecasters. *Journal of Business and Economic Statistics* **25**: (April): 191–200.

Capistran C, Timmermann A. 2009. Forecast combination with entry and exit of experts. *Journal of Business and Economic Statistics* **27**: 429–440.

Davies A, Lahiri K. 1995. A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics* **68**: 205–227.

Davies A, Lahiri K. 1999. Re-examining the rational expectations hypothesis. In *Analysis of Panels and Limited Dependent Variable Models*, Hsiao C, Lahiri K, Lee LF, Pesaran H (eds). Cambridge University Press: Cambridge, UK; 226–254.

Engelberg J, Manski CF, Williams J. 2006. *Comparing the point predictions and subjective probability distributions of professional forecasters*. Working paper W11978, National Bureau of Economic Research, Cambridge, MA.

Engelberg J, Manski CF, Williams J. 2009. Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business and Economic Statistics* **165**: 146–158.

Fair RC, Shiller RJ. 1989. The informational content of ex ante forecasts. *Review of Economics and Statistics* **71**(2): 325–331.

Giordani P, Soderlind P. 2003. Inflation forecast uncertainty. *European Economic Review* **47**: 1037–1059.

Hafer RW, Hein SE. 1985. On the accuracy of time-series, interest rate, and survey forecasts of inflation. *Journal of Business* **58**: 377–398.

Horowitz J, Manski CF. 1998. Censoring of outcomes and regressors due to survey nonresponse: identification and estimation using weights and imputations. *Journal of Econometrics* **84**(1): 37–58.

Keane MF, Runkle DE. 1990. Testing the rationality of price forecasts: new evidence from panel data. *American Economic Review* **80**: (September 1990): 714–735.

Manski CF. 2003. *Partial Identication of Probability Distributions*. New York: Springer-Verlag.

Manski CF. 2007. *Identification for Prediction and Decision*. Harvard University Press: Cambridge, MA.

Patton A, Timmermann A. 2008. *Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion*. Working paper, Oxford/University of California San Diego.

Pennacchi GG. 1991. Identifying the dynamics of real interest rates and inflation: evidence using survey data. *Review of Financial Studies* **4**: 53–86.

Pesaran MH, Weale M. 2006. Survey expectations. In *Handbook of Economic Forecasting*, Elliott G, Granger CWJ, Timmermann A (eds). Elsevier: Amsterdam; 715–776.

Rich R, Tracy J. 2003. *Modeling uncertainty: predictive accuracy as a proxy for predictive confidence*. Staff reports, no. 161, Federal Reserve Bank of New York.

Romer C, Romer DH. 2000. Federal Reserve Information and the Behavior of Interest Rates. *American Economic Review* **90**(3): 429–457.

Thomas L Jr. 1999. Survey measures of expected U.S. inflation. *Journal of Economic Perspectives* **13**(4): 125–144.

Zarnowitz V, Lambros L. 1987. Consensus and Uncertainty in Economic Prediction. *The Journal of Political Economy* **95**(3): 591–621.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

APPENDIX

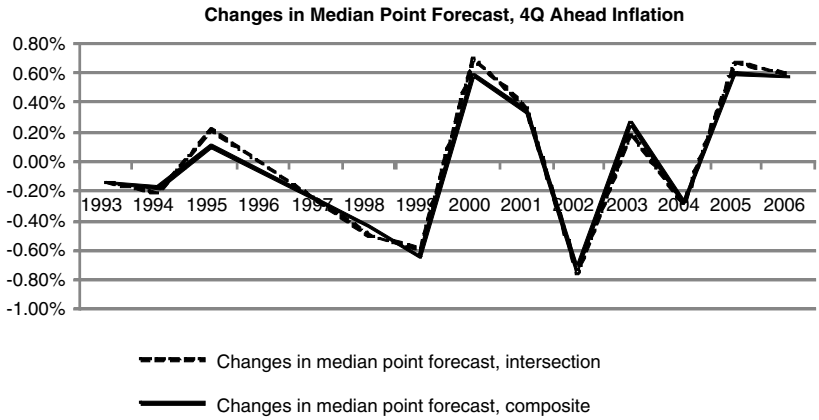

Figure A.1. Compositional effects on changes in median point forecast

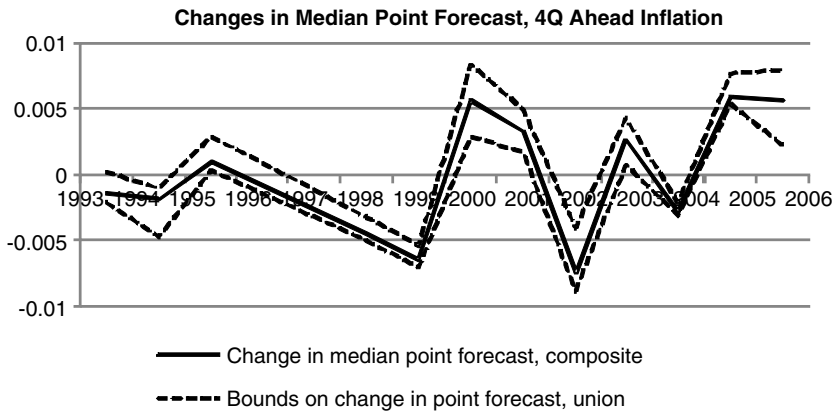

Figure A.2. Bounds on changes in median point forecast

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/jae