

# Inflation Forecast Uncertainty

Paolo Giordani

Paul Söderlind\*

April 17, 2002

## Abstract

We study the inflation uncertainty reported by individual forecasters in the Survey of Professional Forecasters 1969-2001. Three popular measures of uncertainty built from survey data are analyzed in the context of models for forecasting and asset pricing, and improved estimation methods are suggested. Popular time series models are evaluated for their ability to reproduce survey measures of uncertainty. The results show that disagreement is a better proxy of inflation uncertainty than what previous literature has indicated, and that forecasters underestimate inflation uncertainty. We obtain similar results for output growth uncertainty.

**Keywords:** survey data, Survey of Professional Forecasters, GDP growth, VAR, T-GARCH

**JEL:** E31, E37, C53

---

\*Stockholm School of Economics (Giordani); Stockholm School of Economics and CEPR (Söderlind). Correspondance to: Paul Söderlind, Stockholm School of Economics, PO Box 6501, SE-113 83 Stockholm, Sweden; E-mail: Paul.Soderlind@hhs.se. We thank Tom Stark for help with the data; Kajal Lahiri, Lars-Erik Öller, and two anonymous referees for comments.

# 1 Introduction

Modern economic theory predicts that agents' behavior depends on their assessment of the probabilistic distribution of future economic data. It is only under very restrictive assumptions that the point forecast is sufficient to characterize their choices. In general, higher moments also matter. This paper focuses on *inflation uncertainty* as measured by the Survey of Professional Forecasters (SPF) since the late 1960s. It also studies the *real GDP growth uncertainty* from the same survey, which is only available since the early 1980s.

The most common way to assess forecast uncertainty is by estimating some kind of time series model. There are several situations when survey data on expectations/uncertainty are preferable to time series models, for instance, when

- a series has recently undergone a structural change, for example, the adoption of an inflation target;
- different time series methods disagree and it is difficult to point out the best method;
- whenever an empirical rather than a normative measure of uncertainty is needed, so that the interest focuses on actual agents' expectations.

As an example, consider Sargent's (1993) claim that a policy of reducing the inflation rate need not cause any output loss—provided the change in regime is credible. To be made operational, the claim needs a measure of credibility. One way to assess credibility is then to consider the mean and width of agents' forecast error bands. As another example, forecast error bands make it possible to evaluate the credibility of inflation targets, including the tolerance intervals, used by many central banks. In such circumstances a survey measure of uncertainty has clear advantages over an econometric estimate. If a change in regime is suspected, these advantages are magnified.

But even if having a measure of uncertainty from survey data is often desirable, there is no clear, uncontroversial, way of extracting such a measure. A main concern of the paper is to show the conceptual and practical importance of how the survey data is used.

The first issue we discuss is how to think about inflation and GDP growth uncertainty when every forecaster reports his own perceived uncertainty, but also disagrees with other forecasters on the point forecast. We use a simple theoretical framework to highlight that

the relevant definition of uncertainty depends on its intended use. For example, we maintain that previous findings (Diebold, Tay, and Wallis, 1999) that forecasters overestimate inflation uncertainty are based on an inappropriate definition of uncertainty, and that the conclusion ought to be reversed.

The second issue we discuss is how uncertainty can be estimated from the individual answers. Using improved (more robust) estimation techniques, we conclude that disagreement on the point forecast, a readily available but (at present) theoretically unfounded measure of uncertainty, is a better proxy for more theoretically appealing measures than previously thought (Zarnowitz and Lambros, 1987). We also show that recent forecasting errors have stronger effects on perceived uncertainty (as in an ARCH model) than found in previous studies (Ivanova and Lahiri, 2000), and that a whole range of different time series models all fail to keep up with regime changes in U.S. inflation uncertainty (especially in the early 1980s).

The rest of the paper proceeds as follows. Section 2 presents the data in the Survey of Professional Forecasters. Section 3 discusses alternative measures of uncertainty from the survey data. Section 4 discusses the estimation of uncertainty. Section 5 presents the empirical results, and Section 6 concludes.

## **2 The Survey of Professional Forecasters**

The data used in this paper are from the Survey of Professional Forecasters (SPF), which is a quarterly survey of forecasters' views on key economic variables. The respondents, who supply anonymous answers, are professional forecasters from the business and financial community. The survey was started in 1968 by Victor Zarnowitz and others of the American Statistical Association and National Bureau of Economic Research. The number of forecasters was then around 60, but decreased in two major steps in the mid 1970s and mid 1980s to as low as 14 forecasters in 1990. The survey was then taken over by the Federal Reserve Bank of Philadelphia and the number of forecasters stabilized around 30. See Croushore (1993) for details.

There is no guarantee that respondents in a survey give their best (in statistical sense) forecasts. They may simply give nonsense answers, or biased answers due to, for example, strategic considerations. For instance, Laster, Bennett, and Geoum (1999) argue that forecasters may have an incentive to publish forecasts that stand out. However, their

argument relies on the answers being public, and the answers to the SPF survey are anonymous. As for the risk of nonsense answers, we believe that the appointment procedure of the respondents, which includes a screening of the candidates, goes far in ensuring that most of the answers, most of the time, accurately represents the respondents' beliefs. We also believe that it is a strength of the SPF that the forecasters are close to important economic decision makers, since this makes it more likely that the survey reflects beliefs that affect the most important pricing and investment decisions. There may still be odd or erroneous (sloppy handwriting...) answers in the SPF data base. This is one of the reasons why we use robust estimation methods which mitigate the problems with outliers.

A unique feature of the SPF is that it asks for probabilities (on top of the usual point forecasts) for a few variables. In particular, since the start in 1968Q4 it asks for probabilities of different intervals of (annual average) GDP deflator inflation, that is, the GDP deflator for year  $t$  divided by the GDP deflator for year  $t - 1$ , minus one. Since 1981Q3 the survey also asks for real output growth (GNP or GDP, growth defined in the same way as for the deflator).<sup>1</sup>

The SPF is a quarterly survey, but we have chosen to focus mostly on the first quarters. The reason is that there are indications that the survey in the other quarters are not comparable across years (at least not before 1981Q3) since the forecasting horizon shifted in a non-systematic way (see Federal Reserve Bank of Philadelphia, 2000). In contrast, the first-quarter data is known to refer to the growth rate of the deflator or output from the previous to the current year (annual average). We later show that we get similar results if we incorporate the “clean” data points (according to Federal Reserve Bank of Philadelphia, 2000) for Q2-Q4.

[Figure 1 about here.]

The inflation and output growth intervals (bins) for which survey respondents are asked to give probabilities have changed over time. There is an open lower interval, a series of interior intervals of equal width, and an open upper interval. The width of the intervals has changed over time (1% before 1981Q3 and after 1991Q4, 2% in the intermediate period) which may influence estimates of variance. We will take this into account.

---

<sup>1</sup>The survey asked about the GNP deflator before 1992, the GDP deflator 1992-1995, and the (chain weighted) GDP price index since 1996. It also asked for nominal GNP growth before 1981Q3, real GNP growth 1981Q3-1991Q4, and real GDP growth since 1992.

The results from this survey are typically reported in three ways: the median point forecast, the dispersion of the individual point forecasts, and the aggregate (or mean) histograms which are constructed by averaging the probabilities from the individual forecasters' histograms. We use these, but also measures of individual (and aggregate) uncertainty estimated from the individual (aggregate) histograms.

As a preview of the data, *Figure 1* shows the aggregate inflation histogram for all first quarters 1969-2001. The means of these “distributions” (vertical lines) follow the well-known story about US inflation.

For most years, the histograms are reasonably symmetric with most of the probability mass in interior intervals, so the SPF data seems useful for eliciting measures of uncertainty. However, 1985 is a striking exception with 60% of the mass in the open lower interval of inflation lower than 4%: it is clear that the placement of the survey intervals did not keep track with the lower inflation after the “Volcker disinflation.” The intervals were adjusted only in 1985Q2. It is very difficult to say anything about the moments of the distribution with so little information as in 1985Q1. Moreover, Federal Reserve Bank of Philadelphia (2000) suspects that the surveys in both 1985Q1 and 1986Q1 may have asked for the wrong forecasting horizon. We will therefore disregard these two observations in the rest of the analysis. (As a robustness check, we substituted 1985Q2 and 1986Q2 for 1985Q1 and 1986Q1, and got very similar results as without these years.)

### 3 Which Measure of Uncertainty?

In this section we discuss what we should take as the relevant measure of inflation or output growth uncertainty. Not surprisingly, the answer depends on what we want to use it for. There are three main candidates: disagreement among forecasters, average individual forecast error variance (or standard deviation), and the variance of SPF's aggregate histogram. They have all been used in previous research and macroeconomic analysis.<sup>2</sup>

Disagreement on the point forecast has the advantages of being readily available and easy to compute. The disadvantages are also clear. This measure becomes meaningless as the number of agents goes to one or when agents have the same information and agree on

---

<sup>2</sup>Studies that have used measures of uncertainty extracted from survey data to analyze the effects of inflation uncertainty on macroeconomic variables include Barnea, Amihud, and Lakonishok (1979), Lahiri, Teigland, and Zaporowski (1988), Levi and Makin (1980), Mullineaux (1980), Bomberger and Frazer (1981), Melvin (1982), Makin (1982, 1983), Ratti (1985), and Holland (1986, 1993).

the model to use in forecasting. In this case, the measure of uncertainty is zero as if the economy was deterministic. One of the questions tackled in this paper is if disagreement mirrors other measures of uncertainty that are theoretically more appealing, but less easily available.

The average standard deviation of individual histograms does not have the drawbacks of the first measure. It is attractive because it is easy to associate with the uncertainty of a representative agent. On the other hand, it sweeps disagreement under the rug.

The aggregate uncertainty is computed as the variance in the aggregate histogram. It incorporates both individual uncertainty and disagreement.

We now set up a small model of many forecasters in order to discuss the relation between these (and one more) measures of uncertainty. This model highlights and extends several important results in previous work by Zarnowitz and Lambros (1987), Lahiri, Teigland, and Zaporowski (1988), Granger and Ramanathan (1984), and others.

In this model, *individual forecasters* face different, but correlated, information sets which they use to make the best possible inflation (or output growth) forecast. By “best” we mean the conditional mean, which minimizes the expected squared forecasting error. For notational convenience, the information set and forecasting model of forecaster  $i$  is summarized by a scalar signal,  $z_i$ . This signal is useful for forecasting only if it is correlated with actual inflation,  $\pi$ . Let  $\text{pdf}(\pi|i)$  be the probability density function of inflation conditional on receiving the signal of forecaster  $i$ —this should correspond to the histogram reported by forecaster  $i$  in a given quarter (to economize on notation we suppress time subscripts, with the hope that the context makes it clear that we are discussing separate distributions for each period). The mean and variance of this distribution are  $\mu_i$  and  $\sigma_i^2$ , which can be different for different forecasters. One natural measure of inflation uncertainty is the average (across forecasters) individual uncertainty, which we denote by  $E(\sigma_i^2)$ . This is the uncertainty a reader of forecasts faces if he randomly picks (and trusts) one of the point forecasts (see, for instance, Batchelor and Dua, 1995).

**Example 1** *As an example, suppose  $\pi$  and  $z_i$  have a multivariate normal distribution with zero means (to simplify the algebra), variances  $s_{\pi\pi}$  and  $s_{ii}$ , and covariance  $s_{\pi i}$ . In this case,  $\text{pdf}(\pi|i)$  is a normal distribution with mean  $\mu_i = (s_{\pi i}/s_{ii}) z_i$  and variance  $\sigma_i^2 = s_{\pi\pi} - s_{\pi i}^2/s_{ii}$ , which is the standard least squares result.*

Anyone who has access to the survey data can use the cross sectional average of the

individual forecasts,  $E \mu_i$ , as a *combined forecast*.<sup>3</sup> If the number of forecasters goes to infinity, then all individual movements in the forecast errors are averaged out, and only the common movements remain. It is straightforward to show (see reference above) that the (expected) forecast error variance of an unweighted average of unbiased forecasts equals the average covariance of individual errors.<sup>4</sup> This must be less than the average individual uncertainty,  $E(\sigma_i^2)$ , so there is a gain from using a combined forecast as long as the individual forecast errors are unbiased and not perfectly correlated. This theoretical argument suggests that the mean forecast in the SPF should be assigned a smaller uncertainty than the average individual uncertainty.

**Example 2** *To continue the previous example, suppose that individual signals have the same variance and covariances. It is straightforward to show that the forecast error variance of the combined forecast then simplifies to  $E(\sigma_i^2) - \text{Var}(\mu_i)$ , which is the individual forecast error variance minus the cross sectional (across forecasters) variance of point forecasts. The combined forecast is thus better than individual forecasts, especially if forecasters disagree.*

As mentioned, the SPF combines the individual histograms into aggregate (or mean) probabilities, by taking the average (across forecasters) probability for each inflation interval. To see how this *aggregate distribution* is related to individual uncertainty and disagreement among forecasters, think of both future inflation and forecaster  $i$ 's signal as random variables. Also, let  $\text{pdf}(i)$  be the density function of receiving the signal of forecaster  $i$ . We then see that the aggregate distribution,  $\text{pdf}_A(\pi)$ , which averages  $\text{pdf}(\pi|i)$  across forecasters, amounts to calculating the “marginal” distribution of  $\pi$

$$\text{pdf}_A(\pi) = \int_{-\infty}^{\infty} \text{pdf}(\pi|i) \text{pdf}(i) di. \quad (1)$$

As before, let  $\mu_i$  and  $\sigma_i^2$  be the mean and variance in forecaster  $i$ 's distribution,

---

<sup>3</sup>It is well established, both in theory and practice, that an unweighted combination of several different methods/forecasters typically reduces the forecast uncertainty. See, for instance, Granger and Ramanathan (1984) for a general discussion of optimal combinations; and Zarnowitz (1967) and Figlewski (1983) for applications to inflation data.

<sup>4</sup>With  $n$  forecasters we have  $\text{Var}(\pi - \sum_{i=1}^n \mu_i/n) = E \sigma_i^2/n + E \gamma_{ij}(1 - 1/n)$ , where  $E \gamma_{ij}$  is the average covariance of two individual forecast errors.

pdf( $\pi|i$ ). We know that the variance of the distribution in (1) is<sup>5</sup>

$$\text{Var}_A(\pi) = E(\sigma_i^2) + \text{Var}(\mu_i), \quad (2)$$

so the variance of the aggregate distribution of  $\pi$  can be decomposed into the average of the forecasters' variances (average individual uncertainty) and the variance of the forecasters' means (disagreement).<sup>6</sup>

It is not obvious what the aggregate distribution represents. It is less “informed” than the individual conditional distributions (higher forecast error variance, see (2)), but it is more informed than the unconditional distribution (when there is no signal at all) of  $\pi$ . To see the latter, note that we get the unconditional distribution by integrating the distribution once more: this time over the distribution of cross-sectional (across forecasters) means—which is essentially the same as integrating across macroeconomic states.

An appealing, but deceptive, interpretation of the aggregate distribution is that it captures the uncertainty a reader of forecasts faces if he randomly picks (and trusts) one of the point forecasts. Such a reader of forecasts may believe that he/she faces two sources of uncertainty: which forecast to trust and then that forecast's uncertainty. This is wrong, however, if the individual forecaster understands that he could make a more precise forecast if he had all the information of the other forecasters. Forecasters who calculate conditional expectations do, so  $E(\sigma_i^2)$  is the correct measure of uncertainty for this reader.

**Example 3** *To continue the previous example, the aggregate distribution becomes a normal distribution with mean  $E\mu_i$  and variance  $\text{Var}_A(\pi) = \sigma_i^2 + \text{Var}(\mu_i)$ . This means that the uncertainty of the aggregate distribution and the uncertainty of the combined forecast are symmetrically placed around average individual uncertainty.*

This simple model of forecasting suggests a few things. First, individual uncertainty seems to be a key measure of uncertainty. Second, it is hard to justify the aggregate distribution from a forecasting perspective—unless we believe that individual forecasters underestimate uncertainty (whether this is the case empirically is discussed in Section 5.5). The aggregate distribution could still be interesting—from an economic perspective. The beliefs of an agent will influence his consumption and investment decisions, so the

---

<sup>5</sup>For any random variables  $y$  and  $x$  we have  $\text{Var}(y) = E\text{Var}(y|x) + \text{Var}E(y|x)$ , if the moments exist.

<sup>6</sup>For a different (and earlier) derivation of this decomposition, see Lahiri, Teigland, and Zaporowski (1988).



aggregate economy is likely to be affected by some kind of average beliefs. For instance, the asset pricing models of Varian (1985) and Benninga and Mayshar (2000) show that if investors have logarithmic utility functions, then asset prices are directly linked to the average (across investors) probabilities. With respect to inflation, this means that the inflation risk premium on a nominal bond would be directly linked to SPF's aggregate distribution (of inflation). Third, care is needed when survey data on point forecasts and forecast uncertainty is evaluated. For instance, it is hard to interpret evidence from studies that surround the consensus point forecast with an error band derived from the aggregate distribution.

## 4 Estimating Uncertainty from Survey Probabilities

This section discusses how we estimate variances from SPF's histograms. It is always tricky to estimate moments from a histogram, but it is even trickier in this case since the width of SPF's inflation intervals has changed over time (1% for most of the sample, but 2% 81Q3-91Q4).

We choose to fit normal distributions to each histogram: the mean and variance are estimated by minimizing the sum of the squared difference between the survey probabilities and the probabilities for the same intervals implied by the normal distribution. This can be thought of as a non-linear least squares approach where the survey probability is the dependent variable and the interval boundaries the regressors. We have also experimented with adding the reported point forecast to the estimation problem, but the results changed very little.

A straightforward, but crude, alternative way to estimate the mean and variance from a histogram is

$$\tilde{E}\pi = \sum_{k=1}^K \bar{\pi}(k) \Pr(k) \text{ and } \tilde{\text{Var}}(\pi) = \sum_{k=1}^K \bar{\pi}(k)^2 \Pr(k) - \tilde{E}\pi^2, \quad (3)$$

where  $\bar{\pi}(k)$  and  $\Pr(k)$  are the midpoint and probability of interval  $k$ , respectively. The lowest and highest intervals, which are open, are typically taken to be closed intervals of the same width as the interior intervals.

The approach in (3) essentially assumes that all probability mass are located at the interval midpoints. An alternative is to assume that the distribution is uniform within each bin (gives same mean estimator, but the variance is increased by 1/12th of the squared

bin width). These assumptions are standard in the literature; see, for instance, Zarnowitz and Lambros (1987), Lahiri and Teigland (1987), and Diebold, Tay, and Wallis (1999). However, the shape of the histograms in *Figure 1*, which often look fairly bell shaped, suggests that this approach overestimates the variance. It rather seems plausible that relatively more of the probability mass within an interval is located closer to the overall mean. This motivates our choice of fitting normal distributions.

[Figure 2 about here.]

*Figure 2.a* shows the aggregate survey probabilities for inflation once again, and *Figure 2.b* shows the difference between the aggregate survey probabilities and the probabilities implied by the fitted normal distributions. The normal distributions (with two parameters) seems to be able to fit most of the intervals (6, 10, or 15 depending on period) most of the time.

[Figure 3 about here.]

*Figure 3* illustrates the effects of fitting normal distributions rather than using the crude estimator (3), by showing the estimated standard deviations of the aggregate distribution of inflation. The crude method produces consistently higher standard deviations with particularly volatile estimates during the 1980s when there were few and wide intervals.<sup>7</sup> While the underlying distribution may well not be normal, we are inclined to believe that normality provides a better approximation than all the mass at bin midpoint or (even worse) uniformity within bins. If this is the case, statistics obtained with the crude method overestimate the variance and is likely to be sensitive to changes in the bin width.<sup>8</sup>

---

<sup>7</sup>An alternative way of adjusting the crude variance, used by Ivanova and Lahiri (2000), is to apply Sheppard's correction; see Kendall and Stuart (1963). It amounts to subtracting  $1/12$  of the squared interval width from the crude variance. For the standard deviations this means approximately 0.04 for most of the sample, and 0.12 for 1982-1991, which would not make much impact.

<sup>8</sup>To assess the sensitivity of the results to the changes in interval widths, we also reestimated the distributions by using 2% intervals throughout the sample. The results are that 2% intervals give only somewhat higher standard deviations—around 0.1 higher. This suggests that the high estimates of the standard deviations during the 1980s is only partly due to the 2% intervals used by SPF at that time. In fact, none of our main results would be affected by using 2% intervals throughout the sample.

## 5 Empirical Results

### 5.1 Time Profile of Inflation and Output Growth Uncertainty

This section describes how U.S. inflation and real output growth uncertainty has changed since the late 1960s.

[Figure 4 about here.]

*Figure 4* shows the average individual standard deviation,  $E(\sigma_i)$ , of inflation and output growth. Inflation uncertainty was low before 1973 and after 1992, but fairly high in between with peaks in the early 1970s, late 1970s, and early 1990s. For instance, a 90% confidence band constructed from a normal distribution would have been  $\pm 1.6\%$  around the point forecast in 1982 and  $\pm 0.8\%$  in 2000. Output growth uncertainty is slightly higher than inflation uncertainty, but the two series are very strongly correlated.

### 5.2 Different Measures of Uncertainty

This section compares different measures of uncertainty. The purpose is to study if the most easily available measure of uncertainty, the degree of disagreement, is a good proxy for theoretically more appealing measures.

*Figure 5.a* shows the aggregate standard deviation of inflation and average individual standard deviation of inflation,  $E(\sigma_i)$ . The aggregate standard deviation is somewhat more volatile than the average individual standard deviation, but the two series are very highly correlated; the correlation coefficient is 0.89.

[Table 1 about here.]

*Figure 5.b* shows the aggregate standard deviation of inflation and a measure of the cross-sectional dispersion of forecasters' point forecasts  $\mu_i$  (disagreement) of inflation. This measure is a quasi-standard deviation, denoted  $qStd(\mu_i)$ , and is calculated as half the distance between the 84th and 16th percentiles of the point forecasts. If these forecasts were normally distributed, then the quasi-standard deviation would coincide with the standard deviation, otherwise it is much more robust to outliers. The figure shows that the aggregate standard deviation and disagreement are about equally volatile and typically move in the same direction; the correlation coefficient is 0.83.

[Figure 5 about here.]

By comparing Figure 5.a and 5.b we see that disagreement and individual inflation uncertainty typically move together: the correlation in *Table 1* is 0.60. Individual inflation uncertainty is higher than inflation disagreement, but the latter is more volatile. Since they add up to aggregate inflation uncertainty (if squared, see (2)), we see that the level of aggregate uncertainty is mostly due to individual uncertainty, but that the fluctuations in aggregate uncertainty are primarily due to fluctuations in disagreement.

We get similar results for output growth, although based on a smaller sample. For instance, the correlation of disagreement and individual uncertainty is 0.44 for output growth and 0.50 for inflation on the same sample (1982-2001). Similarly, the correlation of disagreement and aggregate standard deviation is 0.75 for both output growth inflation.

Our conclusion is that disagreement is a fairly good proxy for other measures of uncertainty that are more theoretically appealing, but less easily available. Bomberger (1996) reaches the same conclusion based on a comparison of disagreement and ARCH measures of the conditional variance.<sup>9</sup> However, previous research on SPF data has found a weaker correlation between disagreement and other measures of uncertainty (for example, Zarnowitz and Lambros, 1987). Besides the fact that we use a longer sample, the difference is in part driven by our choice of fitting normal distributions and of using a robust measure of disagreement: if the traditional methods are used, the correlations reported above are sizably lower, going from 0.83 to 0.68 and from 0.60 to 0.42. To illustrate this, *Figure 5.c* shows that the traditional measure of disagreement (a standard deviation) is higher and much more volatile than our robust method in 5.b.

### 5.3 Uncertainty and Macro Data

This section studies how uncertainty is related to recent macroeconomic data on inflation, GDP growth, and the survey's own forecast errors. We use the real-time data available at the Federal Reserve Bank of Philadelphia (see Croushore and Stark, 2001) to get as close as possible to the information set at the time of forecast.

*Table 2* shows correlations and regression results for average individual inflation uncertainty,  $E(\sigma_i)$ , but similar results hold for the other measures of uncertainty. The last column shows that  $E(\sigma_i)$  is strongly positively correlated with the consensus forecast of infla-

---

<sup>9</sup>Bomberger's (1996) results are discussed in Rich and Butler (1998) and Bomberger (1999).

tion and recent inflation, and strongly negatively correlated with recent GDP growth. This is consistent with previous findings based on other measures of inflation uncertainty.<sup>10</sup> SPF uncertainty is also positively correlated with recent uncertainty and with the absolute value and the positive part of recent forecasting errors, which suggests “ARCH/GARCH” features.

The remaining columns show multiple regressions where  $E(\sigma_i)$  is regressed on different variables. The first two regressions (columns) show that ARCH and GARCH type models of SPF inflation uncertainty indeed work: lags of  $E(\sigma_i)$  itself and the magnitude of lagged forecast errors have significantly positive effects on  $E(\sigma_i)$ . This is different from the findings of Ivanova and Lahiri (2000). One reason may be that they essentially estimate uncertainty by the more volatile crude method (our ARCH regression indeed turns insignificant if we use that method).

The third regression shows that recent GDP growth has no independent explanatory power for SPF inflation uncertainty (the regression coefficient is far from significant), but that inflation has. However, the size of the inflation effect is not huge: a 3 percentage point increase in inflation is associated with an increase of 0.12 in SPF’s average individual standard deviation (which at the end of the sample is around 0.6). The last regression shows that both the ARCH/GARCH and macro evidence continue to hold when we include both sets of variables in the regression. However, the absolute value of recent forecasting errors is no longer significant at the 5% level (it is at the 6% level, though).<sup>11</sup>

[Table 2 about here.]

It is also interesting to see if the positive correlation of SPF inflation uncertainty and inflation forecast holds also on the “micro level” in the sense that forecasters with high point forecasts (relative to the median that year) are more uncertain. This correlation fluctuates substantially over time (in a non-systematic way), and is on average close to zero (0.12). We get similar results for forecasters with extreme point forecasts (far from the median that year, in either direction).

Uncertainty about real output growth seems to be less correlated with ARCH effects and macro data (even controlling for the sample). However, output growth uncertainty

---

<sup>10</sup>See, among others, Levi and Makin (1980), Makin (1982), Mullineaux (1980), Grier and Perry (2000), Evans and Wachtel (1993), and Zarnowitz and Lambros (1987).

<sup>11</sup>Both inflation variables give very similar results, but including both of them in the same regression results in high multicollinearity (low significance). The same is true for the two “ARCH” variables.

is certainly autocorrelated (with an autocorrelation coefficient of 0.5) and high recent inflation seems to increase output uncertainty (significant on the 8% level) while recent GDP growth does not seem to have any effect.

The use of real-time data is not particularly important for the results in this section—revised data gives very similar results. This will change in the next section, however.

## 5.4 Comparison with Time Series Measures of Uncertainty

This section compares the survey uncertainty with the forecast uncertainty of some popular time series models.

We have argued that it sometimes makes sense to use a measure of uncertainty from survey data. Unfortunately, survey data is not always available, and it is then tempting to use a measure of uncertainty estimated with time series techniques instead. Whether some time series methods mimic the expectations of real agents and, if so, which methods come closer, is then a question of interest. We would also like to know whether the time series models that come close to the SPF uncertainty are also supported by data. Starting from a small group of often applied models (homoskedastic VAR, GARCH, asymmetric GARCH, conditional variance as a function of past inflation), we would like to know whether the models that approximate the SPF uncertainty well also are the models that we would select on the basis of standard econometric criteria.

We choose the average individual uncertainty,  $E(\sigma_i)$ , as a benchmark for comparison, but the main conclusions extend to the other survey measures of uncertainty. This first model is a Vector Autoregression (VAR) with homoskedastic errors. We have strong prior expectations that this cannot be a good model of uncertainty, since it is firmly established that inflation errors are heteroskedastic, and the survey data clearly show that uncertainty is positively correlated with the inflation forecast. Nevertheless, since VARs are widely used for forecasting, it is instructive to see how bad the assumption of homoskedastic errors is in practice. We estimate a VAR(3) on quarterly US real-time data (1955Q1-) on GDP deflator inflation, log real GDP, the federal funds rate, and a 3-year interest rate.

The VAR is first estimated using real-time information available to forecasters at the time of submitting their predictions for 1969 inflation to SPF. A standard deviation for the forecast error of inflation is produced (inflation is defined as average deflator during 1969 over average deflator during 1968 minus one, as in the SPF). The VAR is then re-estimated with data available in early 1970 (a “recursive” VAR) and so on. The standard

deviation of the VAR forecast changes over time because more and more data is used in the estimation and because old data is revised. This recursive estimation procedure on real-time data is adopted for all the models that follow, in an attempt to reproduce the information structure available to forecasters.

[Figure 6 about here.]

*Figure 6.a* shows the VAR series of inflation standard deviations together with the average individual inflation uncertainty from the survey,  $E(\sigma_i)$ . The VAR uncertainty series is much smoother than the SPF uncertainty. The correlation of the two series is only 0.23. This result is not surprising, since in a VAR all residuals have equal weights in forming the standard error of the forecast. A quick fix for this poor performance is to estimate the model on subsample that looks more like the current one. We therefore, estimate a “windowed” VAR recursively on the latest fourteen years of data, so the sample size is fixed rather than progressively larger. In our case this gives rather good results, as illustrated in *6.d* the correlation with the SPF uncertainty is 0.72, though the average level is clearly too low.

GARCH models are strong candidates for modelling inflation uncertainty. In fact, both ARCH (Engle, 1982) and GARCH (Bollerslev, 1986) were first applied to quarterly inflation. Encouraged by the signs of GARCH effects in Section 5.3, we try several types of GARCH models, starting with a standard GARCH(1,1).

There are reasons to believe that the standard GARCH model is not a good model of inflation uncertainty, since the model implies that uncertainty is uncorrelated with the inflation level and that an unforecasted fall in inflation produces as much uncertainty as an unforecasted rise. In any case, we try the model since it is widely used.

We estimate a GARCH(1,1) for GDP deflator inflation in the same recursive fashion as the VAR. The mean is modelled as an AR(4) and the sample is the same as in the VAR. There are now three reasons for why the model standard deviation changes over time: as for the VAR more and more data is used and data is revised (parameters estimates change), but more importantly, the GARCH model produces time-varying standard deviations due to lagged forecast errors. *Figure 6.b* shows the result. The time profile is quite different from the SPF uncertainty (the correlation is 0.41). In particular, the GARCH uncertainty fails to capture the increase in inflation uncertainty around the second oil price shock and

the Volcker disinflation. The inflation uncertainty from the GARCH model is also very low on average.

An asymmetric GARCH could potentially solve the problems of the standard GARCH model. We therefore try a T-GARCH(1,1) for inflation (see Zakoian, 1994). The estimation results indicate that the coefficient of the lagged squared error can be safely set to zero, while the coefficient of the asymmetric component is significant and quite sizeable. This characterization of forecast uncertainty is different from that produced by the standard GARCH: positive inflation surprises increase uncertainty, while negative surprises decrease uncertainty. *Figure 6.c* shows the results from a recursive estimation of the T-GARCH model. The implied inflation standard deviation of the forecast error is now more correlated with the SPF uncertainty (the correlation is 0.58 compared to 0.41 for the GARCH model). The average value is also much closer to the average SPF uncertainty. We conclude that T-GARCH measures of uncertainty should be preferred to GARCH measures based on both econometric testing and on their ability to approximate the SPF uncertainty.

Section 5.3 showed that the SPF uncertainty may be almost as strongly related to the inflation level as to ARCH/GARCH effects. This suggest modelling actual inflation volatility in the same way. We therefore regress the squared forecast errors (from an AR(4)) on the average inflation in the previous four quarters. *Figure 6.d* shows the fitted results, transformed into a standard deviation. This series actually mimics the SPF uncertainty better than the GARCH and T-GARCH models do (the correlation is 0.67).

Even if some of the time series models perform better than the others, all of them fail to capture the increase in inflation uncertainty in the early 1980s. Our opinion is that the large inflation uncertainty in the early 1980s reflects a regime shift (the Volcker disinflation) and is therefore unlikely to be picked up by any time series model in which the conditional variance is a function of past data. Indeed, one may see this as a strong argument in favor of using survey measures of uncertainty in period of suspected structural breaks.

The results for real output growth are fairly similar to those for inflation: the VAR model fails badly (the correlation with the survey uncertainty is only 0.15) unless old data is discarded (the “windowed” VAR has a correlation of 0.43), and both the GARCH and T-GARCH models are clear improvements (the correlations with the survey uncertainty are 0.72 and 0.49, respectively), but the traditional heteroskedasticity model gives a very



odd picture (the correlation with the survey uncertainty is -0.5). Unfortunately, the survey uncertainty of real output growth is only available from 1981Q3, so it hard to say anything about how the time series models handle clear regime changes.

The use of real-time data, as compared to revised data, improves the performance of the GARCH, T-GARCH and “traditional heteroskedasticity” models in proxying survey uncertainty.<sup>12</sup> However, the ranking of the models is not affected, and they all have problems in capturing structural breaks.

## 5.5 Do Forecasters Underestimate Uncertainty?

The last main issue we study is if the survey uncertainty is “correct” in the sense of generating measures of uncertainty that correspond roughly to the objective uncertainty. In particular, we are interested in studying if forecasters underestimate the objective uncertainty, as is often claimed. For instance, Thaler (2000) writes

“Ask people for 90 percent confidence limits for the estimates of various general knowledge questions and the answers will lie within the limits less than 70 percent of the time.”

The natural way of approaching this question is to use the survey data to construct confidence intervals around the point forecasts, and then study if the  $x\%$  confidence intervals cover  $x\%$  of the actual outcomes of GDP deflator inflation (as they should in a large sample). Note that the question we ask involves evaluation of unconditional coverage.<sup>13</sup>

To make this operational, we assume that the forecast errors are approximately normally distributed and construct different confidence bands around both individual point forecasts and the consensus point forecast.

[Table 3 about here.]

*Table 3* shows that individual forecasters underestimate uncertainty: the actual GDP deflator inflation falls inside the 90% confidence bands in only 72% of the observations

---

<sup>12</sup>As expected, the VAR results are not much affected, since estimates of uncertainty are less dependent on recent data.

<sup>13</sup>While correct unconditional coverage is necessary for optimality of the forecast distribution, it is not sufficient if the innovation process is not iid, conditional coverage being also relevant in this case. See Christoffersen (1998) and Diebold, Tay, and Wallis (1999) for a distinction between conditional and unconditional coverage.

(indeed close to Thaler’s assertion).<sup>14</sup> It is then not surprising that Christoffersen’s (1998) test of correct “coverage” of the confidence bands is rejected (at the 5% level) for 34% of the forecasters (45% if we focus on forecasters who participated in the survey at least 5 years).

We get similar results for the consensus point forecast. Our theoretical forecasting model in Section 3 suggests that the variance for the consensus point forecast should (under some assumptions) be the average individual variance minus disagreement,  $E(\sigma_i^2) - \text{Var}(\mu_i)$ . Actual inflation falls inside such a 90% confidence band only 60% of the time. The theoretical forecasting model also shows that  $E(\sigma_i^2)$  provides an upper bound on the variance of the consensus point forecast. Only in this extreme case do the confidence bands generate the correct coverage.

The question of whether forecasters underestimate uncertainty is also raised by Diebold, Tay, and Wallis (1999), and they conclude that forecasters have overestimated uncertainty (at least since 1980). In contrast, we find that forecasters underestimate uncertainty: even on the subsample 1982-2000, we find that the coverage ratio of individual 90% confidence bands is less than 80%.

The main explanation for the difference is that Diebold, Tay, and Wallis (1999) use the aggregate histogram, which corresponds to the line in Table 3 where we surround the consensus point forecast by a confidence band based on  $\text{Std}_A(\pi)$  (which actually have a 100% coverage over the subsample 1982-2000). This combines the best forecast with the widest confidence intervals—which ought to give a high coverage. Rather, the simple forecasting model in Section 3 suggests that the consensus forecast should be assigned a much smaller uncertainty. It seems to us that the best way of tackling the question of whether forecasters underestimate uncertainty is to use the individual distributions, as we do in the first row of Table 3.

A simple example illustrates the point. Suppose that the sample consists of two forecasters, and that the true distribution of inflation is uniform between 0 and 2. Also suppose that the first forecaster reports a uniform distribution between 0 and 1, and the second a uniform between 1 and 2—so both forecasters indeed underestimate uncertainty. Using the methodology in Diebold, Tay, and Wallis (1999), we would build the aggregate histogram, which is uniform between 0 and 2, and thus conclude that the survey distribution

---

<sup>14</sup>To account for the fact that the number of forecasters has changed over time, we normalize the number of forecasters to one for each year, so that each year is given the same weight in forming the average in Table 3.

is optimal. However, using our approach we would conclude that only half of the individual forecasts are inside the 90% percent error bands.

The results for real output growth are even stronger than for inflation: actual output growth falls inside the individual 90% bands only 51% of the times (80% for inflation), and it falls inside the 90% confidence band around the consensus point forecast (with “combined” standard deviation) only 29% of the times.

Our results thus suggest that forecasters underestimate the actual inflation uncertainty. There are several caveats, however. First, it is possible that the results are driven by small sample problems, for instance, in the point forecast used as the mid-point of the confidence intervals. We do not believe that this alone can account for the result. It is true that most forecasters missed the inflation surges in 1969 and 1973-1974, but the results go through even if we disregard these episodes. Second, the forecast errors could be very far from normal. Once again, we do not believe that this is strong enough to overturn the results. The evidence of non-normality is mixed, and perhaps more importantly, some of the numbers in the table actually violate Chebyshev’s inequality—so the fraction inside the band is too low (according to the point estimate, at least) for any distribution (with finite variance).<sup>15</sup>

If it is true that forecasters underestimate the true uncertainty, how should we then use the survey data? In many cases, it is probably the actual beliefs of economic agents that matter, for instance, for understanding investment decisions, asset pricing, and price/wage setting (including monetary policy credibility). However, from a pure forecasting perspective, it may be reasonable to adjust the numbers. The last three lines in Table 3 therefore give results for the consensus point forecast, but using the other survey measures of uncertainty. The results indicate that both the aggregate standard deviation,  $\text{Std}_A(\pi)$ , and the mean individual uncertainty,  $E(\sigma_i)$ , work fairly well, while the disagreement across forecasters,  $\text{qStd}(\mu_i)$ , generates too narrow confidence intervals.

## 5.6 Different Quarters

This section demonstrates that the results for the first quarters (which we have reported above) hold also for the other quarters.

---

<sup>15</sup>Chebyshev’s inequality says that  $\Pr(|x - E x| \geq C\sigma) \leq 1/C^2$  where  $\sigma$  is the (finite) standard deviation of  $x$ . For instance, the 90% confidence level in Table 3 uses  $C = 1.645$ , so the probability of being inside the band is at least 0.63. This is violated by the “Combined Std.”

*Figure 7* compares the average individual inflation standard deviation,  $E(\sigma_i)$ , for different quarters. All data points with inconsistent handling of the forecasting horizon in SPF (see Federal Reserve Bank of Philadelphia, 2000) are set to missing values—this affects mostly the fourth quarter. There are still a few data points that appear suspect, for instance 1969Q3.

[Figure 7 about here.]

The forecasters are asked to supply forecasts and histograms for calendar year inflation (average deflator level in year  $t$  divided by the average deflator level in year  $t - 1$ ). This means that the forecasting horizon is shorter for later quarters, so their standard deviations should (on average) be lower—and this is also what the figures show.

It is also clear that the standard deviations are highly correlated across quarters (correlations with Q1 are in parentheses). This shows that the first quarters capture the main movements in uncertainty over time, so we should expect Q2-Q4 to deliver the same kind of results. This is indeed the case; we summarize the main findings below.

The correlations of different measures of uncertainty are fairly similar across quarters. For instance, the correlation between disagreement and individual uncertainty is 0.60 for Q1 (see Table 1), 0.68 for Q2, and 0.54 for Q3 and 0.46 for Q4.

The correlations with “ARCH proxies” are also similar across quarters, but later quarters have somewhat lower correlations with inflation and GDP growth. For instance, the correlation of individual uncertainty and lagged inflation is 0.65 for Q1 (see Table 2), but 0.41-0.47 for Q2-Q4.

The correlations with the results from time series models are very similar for Q1-Q3, but Q4 appears to be more correlated with all time series models. This reflects the fact that all these models look more similar to the survey uncertainty in the second half of the sample, which is the period for which survey uncertainty for Q4 is available.

The coverage ratios of individual confidence bands are fairly similar across quarters. For instance, for Q1 the 90% confidence bands cover 72% of the deflator inflation outcomes (see Table 3), for Q2 and Q3 it is around 66%, and for Q4 only 56%.

## 6 Summary

This paper studies the uncertainty about U.S. inflation and real output growth reported by the participants of the Survey of Professional Forecasters 1969-2001. We compare

different measures of uncertainty (average individual uncertainty, disagreement about the point forecast, etc), analyze how uncertainty is related to real-time macro data and time series measures of volatility, and examine if forecasters underestimate actual uncertainty.

Extracting a measure of uncertainty from survey data is not an easy task, however, and several pitfalls need to be discovered and circumvented. We use a simple theoretical forecasting/asset pricing model to help us understand the relation between different measures of uncertainty, and how forecasters perception of uncertainty can be evaluated. We also apply improved estimation techniques which handle the discrete nature of the data (histograms) as well as extreme outliers. Our estimates of inflation uncertainty are therefore much less volatile than in previous studies.

We obtain several interesting results:

- *First*, disagreement on the point forecast is readily available, and therefore often used as an indicator of uncertainty. Although other measures of uncertainty may be more theoretically appealing, we find that our different measures of inflation and output growth uncertainty are highly correlated, so (changes in) disagreement can serve as a reasonable proxy for (changes in) uncertainty.
- *Second*, the survey inflation uncertainty is positively related to recent inflation and inflation forecast errors, and contains a large portion of inertia (autocorrelation).
- *Third*, commonly applied time series models (VAR, GARCH, asymmetric GARCH) have problems with giving even approximately the same time profile of inflation uncertainty as the survey. A VAR re-estimated on recent data only, or a model where inflation volatility is regressed on the recent inflation level actually perform better. However, all these econometric models have problems with capturing the changes in the survey uncertainty around the Volcker disinflation in the early 1980s.
- *Fourth*, forecasters seem to underestimate uncertainty, since actual inflation and output growth fall inside their confidence bands much too seldom.

## **A Appendix: Data**

This appendix presents the data sources.

The GNP/GDP deflator and GDP (chain weighted) price series are from the Bureau of Economic Analysis (available at <http://www.bea.doc.gov/>). The Federal funds rate and the

3-years Treasury Note rate (constant maturity) are aggregated to quarterly from monthly data by taking the average of the data at FRED (available at <http://www.stls.frb.org/fred/data>).

The real-time macro data and the data from the Survey of Professional Forecasters is available from the Federal Reserve Bank of Philadelphia (<http://www.phil.frb.org/>).

## References

- Barnea, A., D. Amihud, and J. Lakonishok, 1979, The effect of price level uncertainty on the determination of nominal interest rates: some empirical evidence, *Southern Economic Journal* pp. 609–614.
- Batchelor, Roy, and Pami Dua, 1995, Forecaster diversity and the benefits of combining forecasts, *Management Science* 41, 68–75.
- Benninga, Simon, and Joram Mayshar, 2000, Heterogeneity and option pricing, *Review of Derivatives Research* 4, 7–27.
- Bollerslev, Tim P., 1986, Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics* 31, 309–328.
- Bomberger, W., and W. J. Frazer, Jr, 1981, Interest rates, uncertainty, and the livingston data, *Journal of Finance* 36, 661–675.
- Bomberger, W. A., 1996, Disagreement as a measure of uncertainty, *Journal of Money, Credit and Banking* 28, 381–392.
- , 1999, Disagreement and uncertainty - reply, *Journal of Money, Credit and Banking* 31, 273–276.
- Christoffersen, P. F., 1998, Evaluating interval forecasts, *International Economic Journal* 39, 841–862.
- Croushore, Dean, 1993, Introducing: the survey of professional forecasters, *Federal Reserve Bank of Philadelphia Business Review* pp. 3–13.
- , and Tom Stark, 2001, A real-time data set for macroeconomists, *Journal Of Econometrics* 105, 111–130.
- Diebold, Francis X., Anthony S. Tay, and Kenneth F. Wallis, 1999, Evaluating density forecasts of inflation: the survey of professional forecasters, in R.F. Engle, and H. White, ed.: *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger* (Oxford University Press).

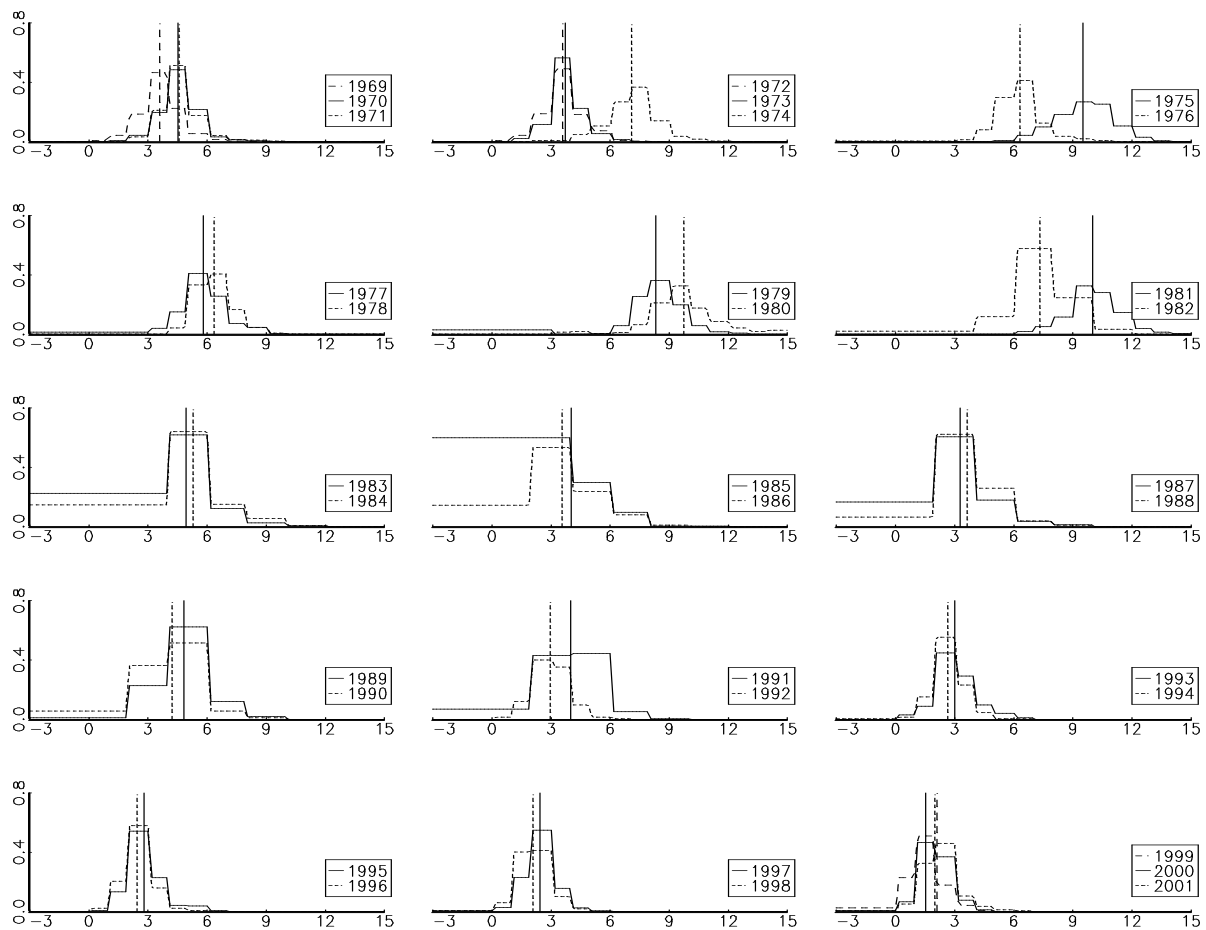
- Engle, Robert F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation, *Econometrica* 50, 987–1008.
- Evans, M., and P. Wachtel, 1993, Inflation regimes and the sources of inflation uncertainty, *Journal of Money, Credit and Banking* 25, 475–511.
- Federal Reserve Bank of Philadelphia, 2000, Survey of professional forecasters: documentation for median forecast data, <http://www.phil.frb.org/econ/spf/spfmeddoc.html>.
- Figlewski, Stephen, 1983, Optimal price forecasting using survey data, *Review of Economics and Statistics* 65, 13–21.
- Granger, Clive J., and Ramu Ramanathan, 1984, Improved methods of combining forecasts, *Journal of Forecasting* 3, 197–204.
- Grier, K., and M. Perry, 2000, The effects of real and nominal uncertainty on inflation and output growth: some garch-m evidence, *Journal of Applied Econometrics* 15, 45–58.
- Holland, S. J., 1986, Wage indexation and the effect of inflation uncertainty on unemployment: an empirical analysis, *American Economic Review* 76, 235–243.
- , 1993, Uncertain effects of money and the link between the inflation rate and inflation uncertainty, *Economic Inquiry* pp. 39–51.
- Ivanova, Detelina, and Kajal Lahiri, 2000, A framework for analyzing subjective probability forecasts and a comparison of inflation forecast uncertainty based on survey data and arch-type models, mimeo, University at Albany - SUNY.
- Kendall, Maurice G., and Allan Stuart, 1963, *The advanced theory of statistics* . , vol. 1 (Charles Griffin and Company Ltd: London) 2nd edn.
- Lahiri, Kajal, and Christie Teigland, 1987, On the normality of probability distributions of inflation and GNP forecasts, *International Journal of Forecasting* 3, 269–279.
- , and Mark Zaporowski, 1988, Interest rates and the subjective probability distribution of inflation forecasts, *Journal of Money, Credit, and Banking* 20, 233–248.
- Laster, David, Paul Bennett, and In Sun Geoum, 1999, Rational bias in macroeconomic forecasts, *Quarterly Journal of Economics* 114, 293–318.



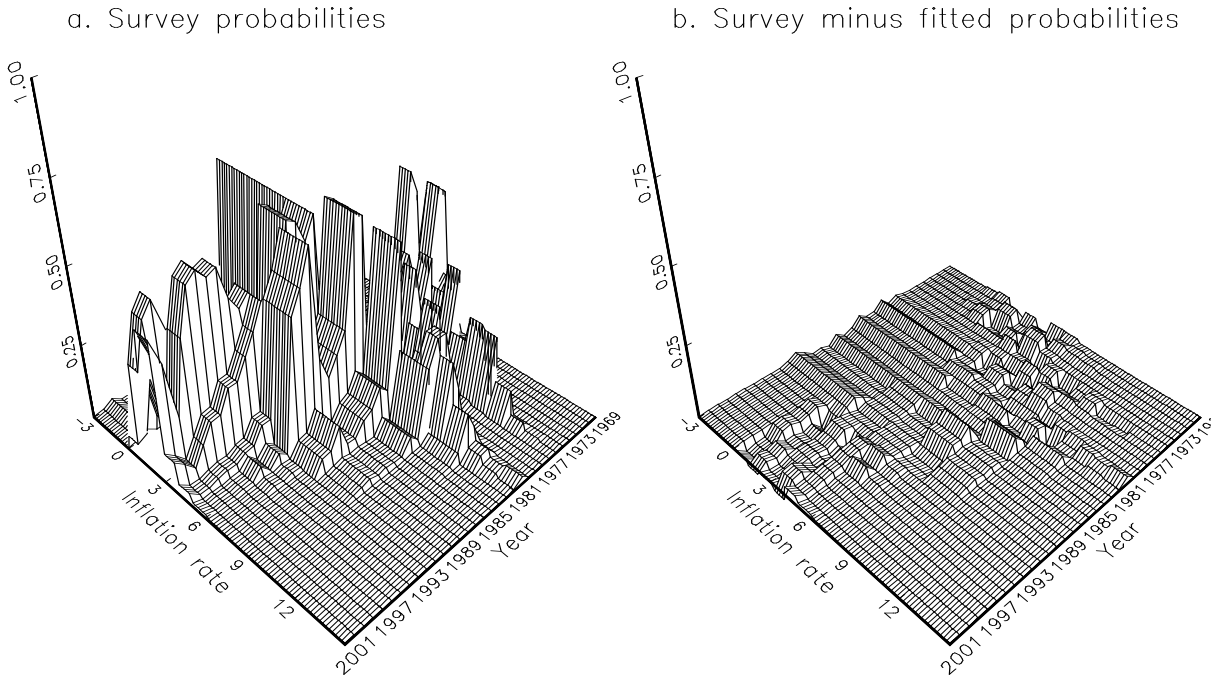
- Levi, M., and J. Makin, 1980, Inflation uncertainty and the phillips curve: some emprirical evidence, *American Economic Review* 70, 1022–1027.
- Makin, J., 1982, Anticipated money, inflation uncertainty, and real economic activity, *Review of Economics and Statistics* 64, 126–134.
- , 1983, Real interest, money surprises, anticipated inflation and fiscal deficits, *Review of Economics and Statistics* 65, 374–384.
- Melvin, M., 1982, Expected inflation, taxation, and interest rates: the delusion of fiscal illusion, *American Economic Review* 72, 841–845.
- Mullineaux, D., 1980, Unemployment, industrial production, and inflation uncertainty in the united states, *Review of Economic and Statistics* 62, 163–169.
- Newey, W. K., and K. D. West, 1987, A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.
- Ratti, R., 1985, The effects of inflation surprises and uncertainty on real wages, *Review of Economic and Statistics* 67, 309–314.
- Rich, R. W., and J. S. Butler, 1998, Disagreement as a measure of uncertainty: a comment on bomberger, *Journal of Money, Credit and Banking* 30, 411–419.
- Sargent, Thomas J., 1993, *Rational expectations and inflation* (Harper Collins).
- Thaler, Richard H., 2000, From homo economicus to homo sapiens, *Journal of Economic Perspectives* 14, 133–141.
- Varian, Hal R., 1985, Divergence of opinion in complete markets: a note, *Journal of Finance* 40, 309–317.
- Zakoian, J. M., 1994, Threshold heteroskedastic models, *Journal of Economic Dynamics and Control* 18, 931–955.
- Zarnowitz, Victor, 1967, *An appraisal of short-term economic forecasts* (NBER: New York).
- , and Louis A. Lambros, 1987, Consensus and uncertainty in economic prediction, *Journal of Political Economy* 95, 591–620.

## List of Figures

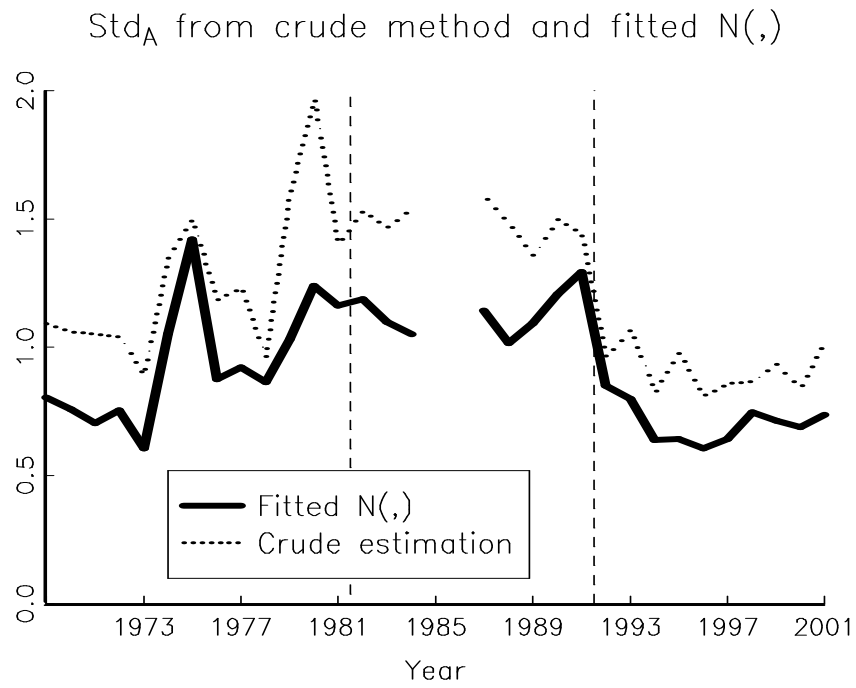
- 1    **Aggregate inflation probabilities in SPF 1969-2001.** This figure shows the aggregate probabilities
- 2    **Aggregate inflation probabilities and estimation error.** Subfigure a shows the aggregate survey pr
- 3    **Aggregate inflation standard deviation estimated in two different ways.** The figure compares stan
- 4    **Individual standard deviations of inflation and real output growth.** The figure compares the aver
- 5    **Different measures of inflation uncertainty.** Subfigure a compares the standard deviations of the ag
- 6    **Inflation uncertainty in survey and time series models.** Subfigure a compares the forecasts error s
- 7    **Individual inflation uncertainty in different quarters.** These figures compares average individual



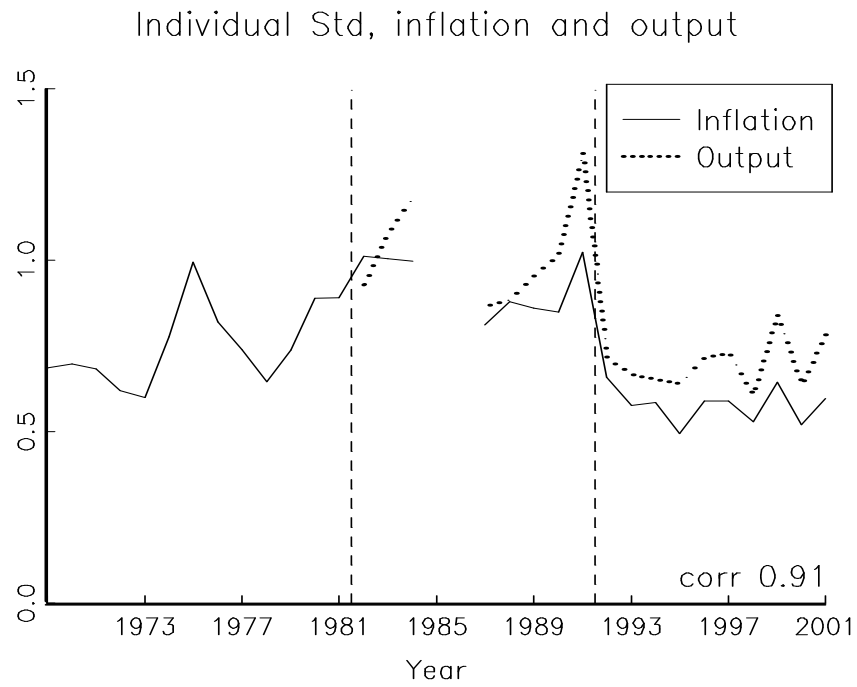
**Figure 1: Aggregate inflation probabilities in SPF 1969-2001.** This figure shows the aggregate probabilities (vertical axis) for different inflation rates (horizontal axis). Estimated means are indicated with vertical lines. Each subfigure shows the probabilities for two (or three) years.



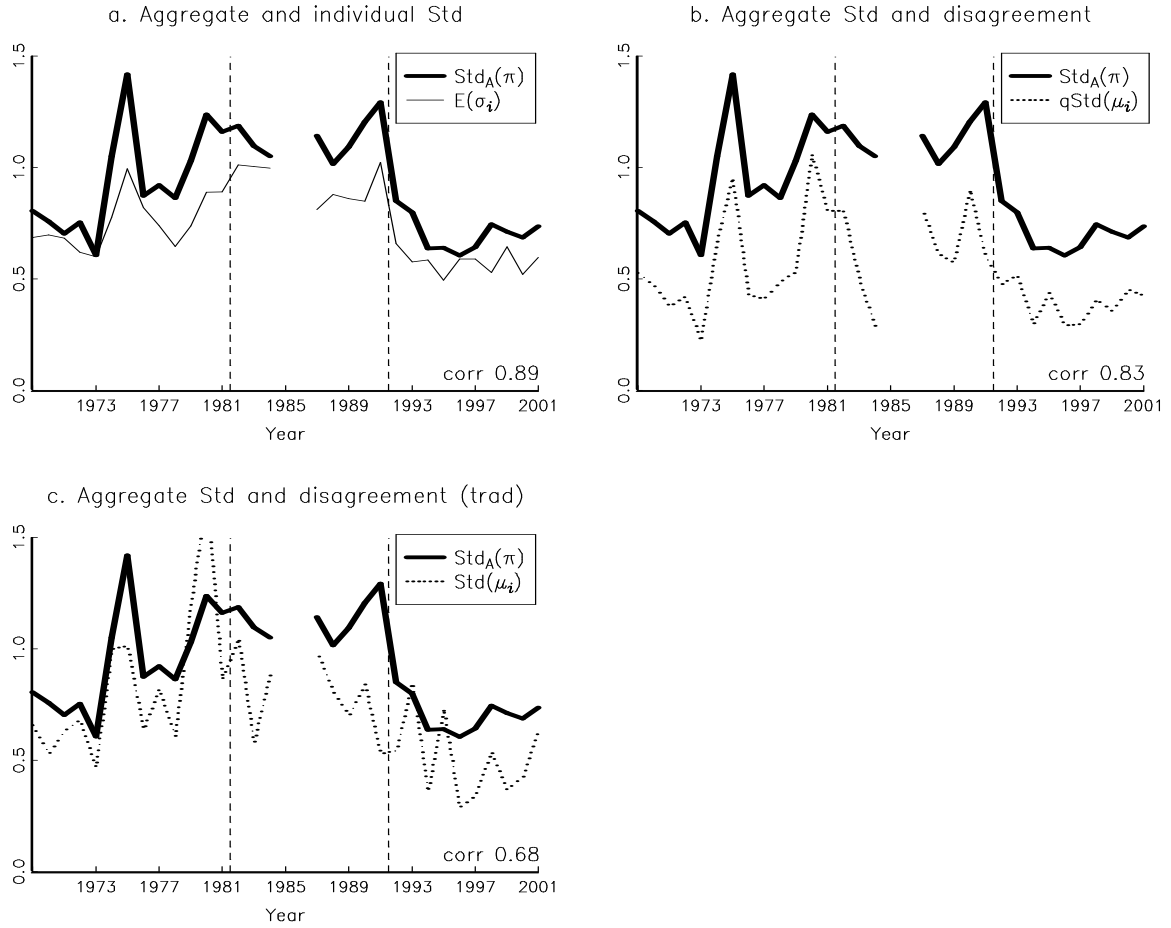
**Figure 2: Aggregate inflation probabilities and estimation error.** Subfigure a shows the aggregate survey probabilities for every year 1969-2001. Subfigure b shows the difference between the survey probabilities and the implied probabilities from fitted normal distributions.



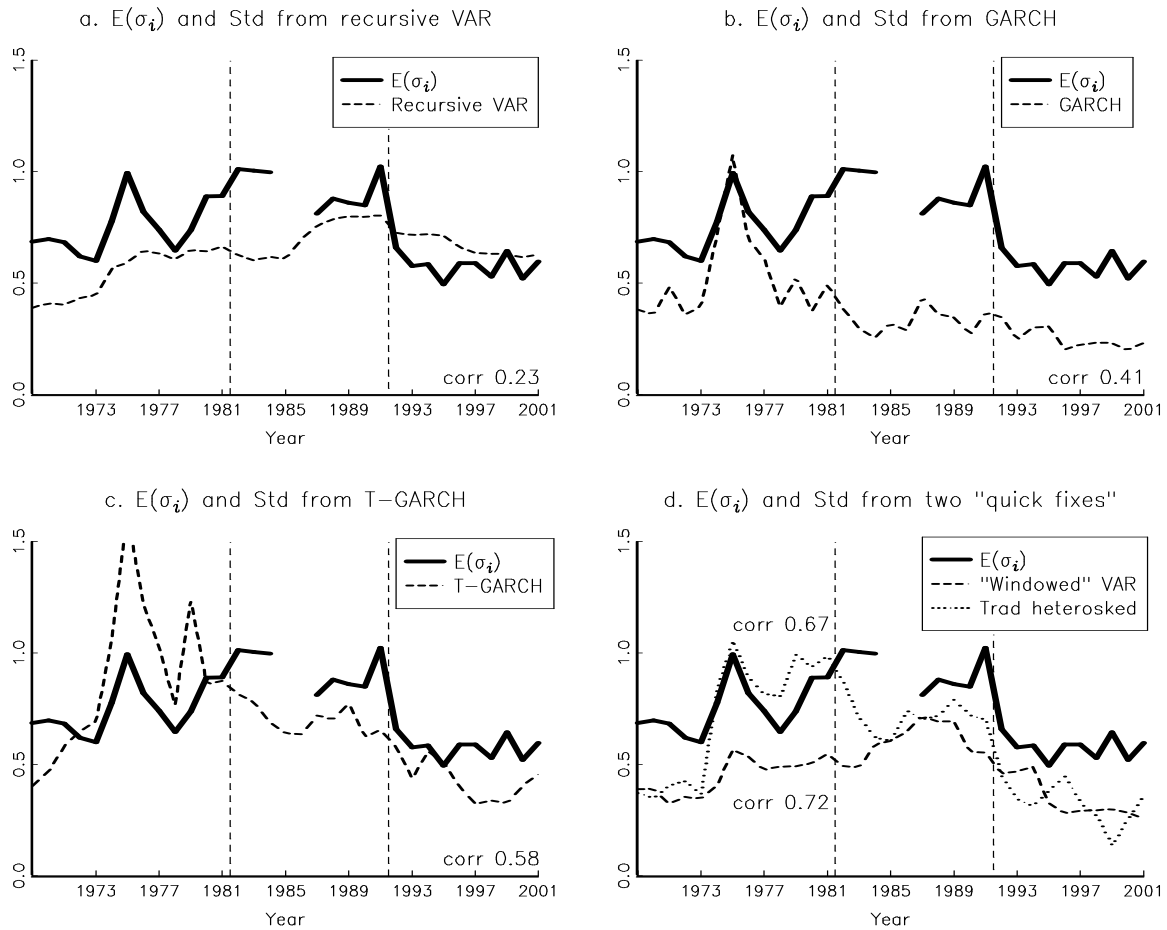
**Figure 3: Aggregate inflation standard deviation estimated in two different ways.** The figure compares standard deviations obtained by fitting normal distribution with those from the crude method (3).



**Figure 4: Individual standard deviations of inflation and real output growth.** The figure compares the average individual standard deviations of inflation (1969-2001) and real output growth (1982-2001).

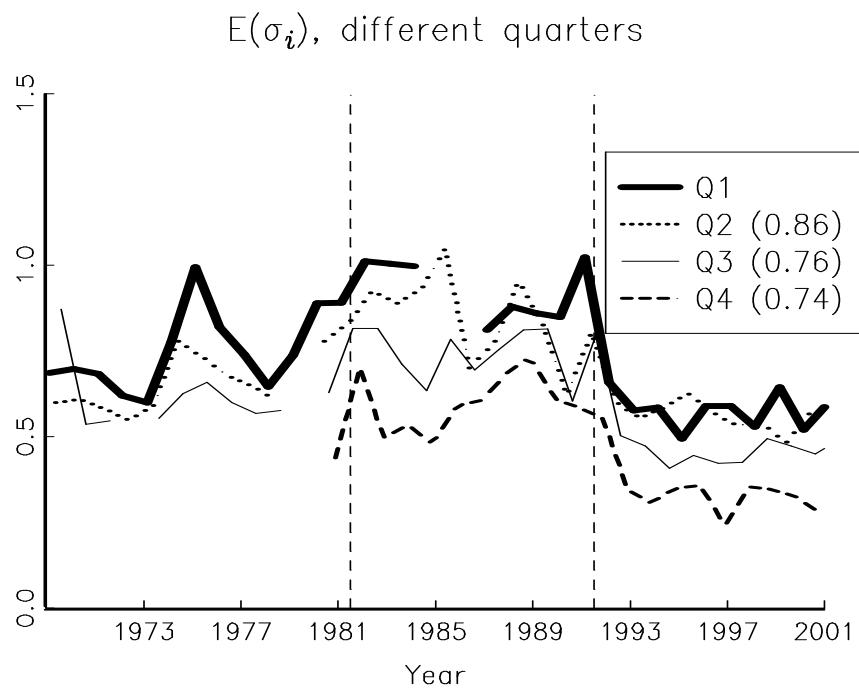


**Figure 5: Different measures of inflation uncertainty.** Subfigure a compares the standard deviations of the aggregate distribution with the average individual standard deviation. Subfigure b compares the standard deviation of the aggregate distribution with the quasi-standard deviation of individual point forecasts. Subfigure c shows a traditional standard deviation of individual point forecasts instead of a quasi-standard deviation.



**Figure 6: Inflation uncertainty in survey and time series models.** Subfigure a compares the forecasts error standard deviations from a VAR model estimated on a longer and longer sample with the aggregate standard deviation from SFP. Subfigure b compares with the average individual standard deviation instead.





**Figure 7: Individual inflation uncertainty in different quarters.** These figures compares average individual inflation uncertainty in different quarters. The forecast is for the calendar year inflation (annual average). The numbers in parentheses (legend) are correlations with Q1.

## List of Tables

1	Correlations of different measures of inflation uncertainty in SPF . .	35
2	Correlation of inflation uncertainty and real-time macro data . . . .	36
3	Comparison of confidence bands and ex post inflation . . . . .	37

**Table 1: Correlations of different measures of inflation uncertainty in SPF**

	$\text{Std}_A(\pi)$	$E(\sigma_i)$	$\text{qStd}(\mu_i)$	$\frac{\text{qStd}(\mu_i)}{\text{Std}_A(\pi)}$
$\text{Std}_A(\pi)$	1.00			
$E(\sigma_i)$	0.89*	1.00		
$\text{qStd}(\mu_i)$	0.83*	0.60*	1.00	
$\text{qStd}(\mu_i)/\text{Std}_A(\pi)$	0.40*	0.02	0.79*	1.00

A star \* denotes significantly different from zero on the 5% level, using a Newey and West (1987) GMM test. The sample is 1969-2001, excluding 1985 and 1986.  $\text{Std}_A(\pi)$  is the standard deviation of SPF's aggregate distribution;  $E(\sigma_i)$  is the cross-sectional average individual standard deviation;  $\text{qStd}(\mu_i)$  is a cross-sectional quasi-standard deviation of individual point forecasts.

**Table 2: Correlation of inflation uncertainty and real-time macro data**

	Multiple regressions of $E(\sigma_i)$ , columnwise				Correlation with $E(\sigma_i)$
Consensus point forecast					0.62*
Inflation $_{t-1}$			0.04*	0.02*	0.65*
GDP growth $_{t-1}$			-0.01	0.01	-0.45*
Lag of $E(\sigma_i)$		0.73*		0.63*	0.73*
Abs(forecast error $_{t-1}$ )	0.10*	0.08*		0.04	0.39*
max(forecast error $_{t-1}$ , 0)					0.24
$R^2$	0.15	0.63	0.44	0.67	—

A star \* denotes significantly different from zero on the 5% level (see Table 1). All regressors are lagged one quarter. The sample is 1969-2001, excluding 1985 and 1986. The consensus point forecast is the average individual point forecast in SPF; Inflation and GDP growth are year-over-year (real-time data, as available when forecast is made); the own lag is the lagged value (one year) of the variables in the column headings; the forecast error is the actual GDP deflator (real-time data, as available after the end of the forecasting horizon) inflation minus the consensus forecast.

**Table 3: Comparison of confidence bands and ex post inflation**

Type of confidence band	Confidence level ( $x\%$ ):		
	90%	80%	66%
Around individual point forecasts, with Std from:			
$\sigma_i$	0.72 <sup>†</sup>	0.62 <sup>†</sup>	0.48 <sup>†</sup>
Around consensus point forecasts, with Std from:			
“Combined Std” $\sqrt{E(\sigma_i^2) - \text{Var}(\mu_i)}$	0.60*	0.43*	0.23*
Std <sub>A</sub> ( $\pi$ )	0.90	0.90	0.73
E( $\sigma_i$ )	0.90	0.80	0.63
qStd( $\mu_i$ )	0.57*	0.47*	0.37*

This table shows the fraction of years when actual inflation is inside the  $x\%$  confidence bands. A star \* denotes significantly different from the nominal confidence level on the 5% level, using Christoffersen's (1998) test. A dagger † denotes that Christoffersen's (1998) test is done for each individual forecaster: the null hypothesis is rejected at the 5% level for 34%, 21%, and 23% of the forecasters, respectively. The confidence bands are calculated assuming a normal distribution and are calculated as: mean inflation forecast  $\pm$  the critical value times the standard deviation. Actual inflation is measured as the percentage change in the GDP deflator (annual-average, revised data). The sample is 1969-2000, excluding 1985 and 1986.