WHAT CAN PROBABILISTIC FORECASTS TELL US ABOUT INFLATION RISKS?*

Juan Angel García†
Andrés Manzanares‡

First version: September 2006
This version: April 2007

Abstract

Inflation expectations are a crucial element in economic theory and monetary policy. As such, their developments are nowadays closely monitored by researchers and policymakers alike. Yet, the analysis is often restricted to single point estimates. This paper presents a new methodology to extract measures of inflation risks from the probability distributions of the Survey of Professional Forecasters (SPF) and provides some stylised facts about their dynamics. We also illustrate how a combined analysis of movements in *all three* first moments of inflation expectations may shed new light on some puzzling episodes in the US bond market.

Keywords: *Inflation risk, inflation expectations, Survey of Professional Forecasters (SPF), skew-normal distribution, power divergence estimators*

*We are very grateful to Geert Bekaert, Gonzalo Camba-Mendez, Refet Gürkaynak, Gabriel Perez-Quirós and Thomas Werner for useful comments and suggestions on earlier drafts. We are particularly indebted to Paul Söderlind for many fruitful discussions in the early stages of this project and very useful comments on earlier drafts. Excellent research assistance by Renate Dreiskena is also gratefully acknowledged. Any remaining errors are our responsibility. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the European Central Bank.

†Capital Markets and Financial Structure Division, DG-E, European Central Bank. Email: juan_angel.garcia@ecb.int.

‡Risk Management Division, DG-H, European Central Bank. Email: andres.manzanares@ecb.int.

"...we must consider not only what appears to be the most likely outcome, but also the risks to that outlook..."

– Testimony of Fed Chairman Ben Bernanke to U.S. Senate, July 19, 2006 –

1 Introduction

Inflation expectations are a crucial element in economic theory and monetary policy. As such, their developments are nowadays closely monitored by researchers and policymakers alike. Yet, attention is often restricted to point estimates, while forecasts are, by nature, probabilistic statements. A point forecast is sufficient to characterize optimal economic decisions, and thereby aggregate outcomes and asset prices, only under very restrictive assumptions. In general, the risks surrounding the baseline scenario of any forecast also matter, and monetary policy decisions should take them into account.

There seems to be two main reasons why attention is often restricted to point estimates. First, the lack of quantitative evidence on the risks surrounding inflation expectations.[1] Unfortunately, the overwhelming majority of surveys of inflation expectations only request a single number, making it impossible to gauge the risks surrounding the forecasts. The second problem, a somewhat natural consequence of the lack of reliable measures of inflation risks, appears to be some reservations regarding their information content beyond that of the point estimates to explain key macroeconomic phenomena.[2]

This paper is a first step towards overcoming those two limitations. First, we seek to provide reliable quantitative evidence on the inflation risks surrounding subjective probabilistic forecasts. The Survey of Professional Forecasters (SPF), currently conducted by the Federal Reserve Bank of Philadelphia, offers the possibility to obtain information about the higher-order moments of inflation forecasts because survey participants are asked to assign probabilities to the likelihood of future inflation falling into pre-specified ranges, i.e. a density forecast in the form of a histogram. Extracting accurate measures of risks from those histograms has been the subject of significant research since the pioneering work by Zarnowitz and Llambros [1987] and Lahiri *et al.* [1988] but has proved quite challenging.[3] This paper provides a new methodology to analyse those probabilistic forecasts and documents key stylised facts about their dynamics over the last four decades.

To accurately measure inflation risks, this paper departs from existing literature in several aspects. We interpret the SPF histograms as a realisation of a multinomial distribution through which a (true) underlying density forecast is discretized. The reported histograms then reflect how many of the "*draws*" taken from the underlying subjective distribution lie within each of the intervals set up in the questionnaire. This naturally lends structure to the statistical problem of recovering the underlying density function from the reported histograms, and that structure provides some crucial insights concerning the

[1] Given that the randomness of inflation agents face will be summarised by statistical moments of the probabilistic forecasts reflecting SPF panellists' subjective beliefs, we find it appropriate to refer to inflation "*risks*" instead of (knightian) inflation "*uncertainty*" (see Machina and Rothchild [2007]).

[2] An additonal general concern about survey data is that participants have little incentives to provide their true expectations. However, the superior forecasting performance of surveys recently documented by Ang *et al.* [2006] casts some doubt on such concerns.

[3] In search for more reliable estimates, Giordani and Söderlind [2003] proposed fitting a normal distribution to the SPF histograms instead of the non-parametric methods of earlier literature. More recently, some alternative approaches to overcome the shortcomings of fitting a normal distribution to the SPF histograms and measure inflation uncertainty have been proposed (see Engelberg et al. [2006], Rich and Tracy [2006] and D'Amico and Orphanides [2006]).

appropriate methodology for the problem at hand. Against this background, we propose a small departure from maximum likelihood estimation that provides the robustness needed to extract reliable measures of inflation risks from the SPF histograms. For instance, that robustness allows us to employ a potentially skewed distribution and "*let the data speak*" about perceived asymmetries in inflation risks. Monte Carlo evidence confirms that these methodological contributions lead to significant accuracy gains in our estimates of the mean forecasts and the uncertainty surrounding them.

We document key stylised facts about perceived inflation risks and their dynamics over the last 40 years. They rose sharply with the oil price crises of the 1970s, but also remained relatively high for most of 1980s, before experiencing a significant decline that still persists. Central tendency measures are strongly correlated with uncertainty measures but little with the measures of asymmetry in risks, and higher inflation and higher inflation volatility have been usually associated with higher inflation uncertainty. The business cycle appears to have a strong influence on inflation risks, but that relationship seems to be forward-looking: inflation risks tend to decline ahead of recessions, are strongly influenced by the perceived uncertainty about real GDP growth and also respond to the perceived probability of recession over the horizon of the inflation forecast. Outside recession periods since the early 1990s the impact of the business cycle on both inflation uncertainty and upside risks seems to have weakened significantly. Importantly, although there is some degree of comovement among the three first moments of inflation forecasts, limiting attention to a point estimate, the mean or any kind of location measure, is not sufficient to characterise all the variation in inflation expectations.

We also investigate the extent to which our evidence on the uncertainty and skewness embodied in inflation expectations may shed new light on some puzzling episodes in bond markets, via the potential relationship between perceived inflation risks and premia embodied in bond yields. In particular, we explore the dynamics of perceived inflation risks during the Volcker disinflation period. Most of existing research has focused on the so-called "inflation scares" faced by the Volcker Fed (see Goodfriend [1993], Goodfriend and King [2005] among others) but a thorough analysis of direct measures of inflation expectations over that period appears missing. This paper aims at filling that void. Our analysis confirms that it took quite long for the Volcker Fed to acquire the desired credibility. It also suggests that changes in perceived inflation risks, particularly upside risks to the inflation outlook, are a fundamental element to understand movements in long-term rates in the 1980s.

The paper is organised as follows. Section 2 presents our methodology to measure the inflation risks embodied in the SPF probabilistic forecasts, in particular our choice of theoretical density function and fitting criterion. Monte Carlo evidence providing quantitative support for the theoretical arguments underpinning our proposed methodology is presented in Section 3, and the main insights from the estimation of the individual and the *consensus* SPF probability distributions are documented in Section 4. The relationship between different moments of inflation forecasts as well as inflation risks and key macroeconomic variables is also reported there. Section 5 focuses on the dynamics of private sector beliefs about inflation during the Volcker disinflation period in the light of our inflation risk measures. Finally Section 6 concludes.

2 A new methodology to analyse the SPF forecasts

The Survey of Professional Forecasters (SPF) gathers information about the subjective probability a number of panelists assign to some macroeconomic variable (inflation and output growth)[4] falling within a number of predetermined intervals of the form $(\alpha_i, \alpha_{i+1}]$, $i = 0, \ldots, I$. Except for the first and the last intervals, which are one-side open (i.e. $\alpha_0 = -\infty$ and $\alpha_I = \infty$), all central intervals in each survey round have equal length.[5] As a result, survey probability distributions, either from individual forecasters or from the aggregation across panelists, are reported in the form of histograms, which reflect the relative probabilities assigned to each interval.

We interpret each of those histograms as a discretized version of a *subjective* density forecast of inflation for a given horizon. A thorough analysis of the information content of the forecast requires the underlying density forecast on which the observed histogram is based to be recovered from the reported frequencies. In probabilistic terms, we can denote each subjective underlying probability as $f_{\pi|F=F_k}$, where F is a discrete random variable that selects a forecaster F_k among the whole panel, $k = 1, \ldots, K$. In turn, the unknown underlying density function $f_{\pi|F=F_k}$ will be assumed to belong to a suitable parametric family of distributions $f_{\pi|F=F_k} = f_{\varrho_k}$ where $\varrho_k \in \Theta \subseteq R^p$ and p is the number of parameters characterising the family distribution. Each survey participant provides a discrete version of that underlying subjective probability distribution. In theory, the survey reply should exactly correspond to the set of integrals of the underlying density function $p_{ik} := \int_{\alpha_i}^{\alpha_{i+1}} f_{\pi|F=F_k}(x)dx$, calculated over each of the intervals $(\alpha_i, \alpha_{i+1}]$ specified in the survey questionnaire. In practice, however, it is unlikely that survey participants compute those integrals to discretise their subjective density forecast and fill in the questionnaire.

Our working assumption is that the probability distributions reported by survey participants are derived as if a number of "random draws" were taken from their underlying subjective density $f_{\pi|F=F_k}$, and the reported figures reflect how many of those draws lie within each interval $(\alpha_i, \alpha_{i+1}]$.[6] Hence, formally, we interpret the set of frequencies provided by the kth forecaster ($\widehat{p}_{ik}$, $i = 0, \ldots, I$) as a summary of the realisations of a multinomial random variable with I classes, whereby the underlying probability for each class can be defined as $p_{ik} = \int_{\alpha_i}^{\alpha_{i+1}} f_{\pi|F=F_k}(x)dx$.

Denoting by $\{Z_{j,k}\}_{j=1,\ldots,n_k}$ the random draws of the discretized multinomial variable underlying the survey response of the kth forecaster, we have:

$$P\left(Z_{1,k} = z_{1,k}, \ldots, Z_{n_k,k} = z_{n_k,k} \mid \widehat{p}_{1,k}, \ldots, \widehat{p}_{I-1,k}\right) = \frac{\prod_{i=0}^{I}(n_k \widehat{p}_{i,k})!}{n_k!}, \quad \forall \left\{z_{j,k}\right\}_{j=1,\ldots,n_k}$$

That expression shows that the set of reported frequencies contains all the information in the sample about the unknown parameter vector ϱ_k. It follows that the frequencies,

[4] In this paper we restrict the analysis to the inflation density forecasts, but our methodology can also be applied to the growth density forecasts. The latter, although also available for a relatively long period, however suffer from the lack of an homogeneous sample as the survey changed from nominal to real growth in the 1980s.

[5] Note that although the length of the intervals in the questionnaire has changed over time, the proposed methodology can easily accommodate those changes.

[6] Obviously, the larger the number of observations drawn, the more the relative frequencies reported approach the values of the integrals described above. Note that in any case the assumption of a concrete number of draws is not needed for the methodology presented here.

$\widehat{p}_{ik}$, are a sufficient statistic for estimating the theoretical probabilities p_{ik} in this simple multinomial framework.[7]

There are a number of important insights that can be obtained from the interpretation of the statistical problem at hand in the context of the (parametric) multinomial framework outlined above. First, the link between the unknown underlying density function $f_{\pi|F=F_k}$ and the observed histograms is formalised. This implies an explicit consideration of the stochastic nature of the statistical problem, which, for example, provides a rationale for the likely presence of rounded probabilities in the SPF data. Second, despite the fact that the multinomial framework does not impose any restrictions on the functional form of the underlying subjective density $f_{\pi|F=F_k}$ it naturally lends structure to the statistical problem at hand, thereby providing some crucial insights that have been overlooked in previous work on this topic. The next sections analyse them in detail.

2.1 The choice of fitting criterion

Our inference problem is to find a parameter vector ϱ characterising a density function such that the resulting vector of interval probabilities is closest to the reported frequencies in the SPF histograms. As the purpose of the exercise is to obtain reliable estimates that help summarise the information content of the SPF distributions, adequate large sample properties of such estimators (consistency, asymptotic normality and asymptotic efficiency[8]) are clearly desirable. In addition, and given the particular features of the SPF probability distributions, robustness in small samples is also a highly desirable property. This section provides a detailed discussion of the statistical considerations behind our approach.

Recent literature that considers the appropriateness of fitting parametric densities to the SPF histograms (Giordani and Söderlind [2003], Rich and Tracy [2006], Engelberg et al.[2006], D'Amico and Orphanides [2006]) has considered least squares as fitting criterion, i.e. minimising the sum of the squared deviations of the observed probabilities with respect to the theoretical ones over the set of intervals. Formally, $LS = \sum_{i=0}^{I} \left(\widehat{p}_i - p_i(\varrho)\right)^2$.

However, minimising the LS criterion (although consistent) is not efficient, for it does not use all available information in the problem at hand. The Appendix elaborates on the shortcomings of (unweighted) least squares in this context, but, intuitively, they stem from the fact that while the LS criterion assigns equal weighting to the fitting errors for each interval, an efficient method would further exploit the fact that SPF data shows a strong bell shape and assign different weights to the fitting errors depending on the probability assigned to them.

Maximum likelihood is a natural alternative in the context of the multinomial framework outlined above. Indeed, for any given histogram the sufficiency of the vector of frequencies $\{\widehat{p}_i\}$, $i = 0, \ldots, I-1$, implies that maximum likelihood would deliver a consistent, asymptotically normal and efficient estimate for the vector ϱ . Specifically the function to be maximised with respect to ϱ would be: [9]

[7] See, for instance, Witting [1985] for a general development on sufficiency for discrete distributions and information theoretical implications.

[8] Consistency, in the sense that the estimator $\widehat{\varrho}$ converges to the true parameter value in probability as the number of observations (draws from the true density forecast in our multinomial framework) $n \to \infty$. Asymptotic efficiency, in the standard sense that no other estimator has smaller variance, as $n \to \infty$.

[9] The likelihood is $\frac{n!}{\prod_{i=0}^{I}(n\widehat{p}_i)!} \prod_{i=0}^{I} p_i^{n\widehat{p}_i}$.

$$\log \text{lik}(\widehat{p}_0, \ldots, \widehat{p}_I, \varrho) \ \propto \ \sum_{i=0}^{I} \widehat{p}_i \log p_i(\varrho) \tag{1}$$

or, equivalently, minimise the log-likelihood ratio G^2 below [10]:

$$G^2 = \sum_{i=0}^{I} \widehat{p}_i \log \widehat{p}_i - \log \text{lik}(\widehat{p}_0, \ldots, \widehat{p}_I, \varrho) \ = \ \sum_{i=0}^{I} \widehat{p}_i \log(\widehat{p}_i / p_i(\varrho)) \tag{2}$$

Applying maximum likelihood estimation to the SPF data is however problematic in practise for two reasons. First, numerical problems stemming from the low probabilities assigned to some of the intervals in the questionnaire, in particular those very close to zero, are likely to arise. Second, in the case of the SPF data, the robustness of the estimator is particularly important given the potential *"contamination"* in the survey responses, which is likely to arise both from model specification (i.e. possible misspecification of the of the theoretical density function assumed) and from "loose" reporting of the discretisation (sloppy handwriting, rounding practices etc) in the SPF histograms. An optimal estimator should therefore exhibit robustness to those features of the SPF data.

Small departures from maximum likelihood estimation in the context of the estimation of multinomial distributions have been extensively studied in the statistical literature. For example, Cressie and Read [1988] propose a family of (efficient) estimators, indexed by the parameter $\tau \in R$), known as "power divergence estimators" (PDE henceforth), that can be defined in general terms as the estimators obtained by minimising with respect to ϱ the following expression:[11]

$$I^{\tau}(\widehat{p}, p) \ = \ \frac{1}{\tau(\tau+1)} \sum_{i=0}^{I} \widehat{p}_i \left[\left(\frac{\widehat{p}_i}{p_i(\varrho)} \right)^{\tau} - 1 \right] \tag{3}$$

To see the close relationship to maximum likelihood consider the limiting case of $\tau = 0$ in equation (3) above

$$I^0 \equiv \lim_{\tau \to 0} I^{\tau} \ = \ \sum_{i=0}^{I} \left[\widehat{p}_i \log \frac{\widehat{p}_i}{p_i(\varrho)} + (\widehat{p}_i - p_i(\varrho)) \right]$$

from which it is immediate that minimizing I^0 with respect to ϱ is equivalent to minimizing the log-likelihood G^2(see equation (2)).[12] Hence maximum likelihood can be interpreted as a limit case of a broader family of criteria.

Cressie and Read [1984] show that, under Birch´s [1964] mild regularity conditions, which are satisfied by our model setup, the power divergence estimators described above satisfy the desired large sample properties, namely consistency, asymptotic efficiency and asymptotic normality delivered by maximum likelihood. Lindsay [1994] proves that, in

[10] The likelihood ratio G^2 is obtained by changing sign and adding the quantity (non dependent on ϱ) $\sum_{i=0}^{I} \widehat{p}_i \log \widehat{p}_i$. In order to allow for a direct comparison to existing results here we follow the notation of Cressie and Read [1988].

[11] The family of power distance estimators encompasses many widely-used estimators such as the Chi-Square criteria of Pearson and Neyman, the Hellinger distance, and the Kullback-Leibler divergence. See Appendix for further details.

[12] Note that it trivially holds that $\sum_{i=0}^{I} \widehat{p}_i - p_i(\varrho) = 0, \ \forall \varrho \in \Theta$.

addition, power divergence estimators are all first-order efficient, the latter being an alternative characterization of the speed of convergence.

However, there are potential pitfalls in transferring those optimal asymptotic properties to small samples, as more robust power distance estimators can perform relatively poorly in terms of efficiency compared to the maximum likelihood estimator when the underlying sample size (in our context the unobservable number of draws carried out to fill in the questionnaire) is small. Our choice of optimal criterion, i.e. the choice of the index parameter τ, therefore also takes into account the small sample properties of the power divergence estimators and the characteristics of the SPF data.

Specifically, an inspection of the SPF data suggests that numerical robustness to *inliers* (i.e. intervals with much lower observed frequency than the theoretical probability obtained from the model density would suggest, for example related to rounding practises), is quite important while outliers are relatively unlikely to be found. For example, about 20% of the respondents assign non-zero probabilities to two bins or less which restricts significantly the information available to the researcher. For such cases, existing results from the comparison of the performance of PDE in the presence of inliers (see Lindsay [1994]) suggest applying a positive, though relatively low, value of the parameter τ. Indeed Cressie and Read [1988] recommend $\tau = 2/3$. Some Monte-Carlo simulations specifically designed to match the particularities of the SPF data appear to confirm that evidence, warranting , among other insights, the use of a positive but rather low τ (see Section 3.2 below).[13]

Taking into account the small sample properties and robustness of the fitting criterion is particularly important in the analysis of the SPF data. Our approach (to be fully specified below), being robust to data inliers, can be uniformly applied across the panel of forecasters. Previous research has instead often been forced to make additional assumptions for the estimation of the densities of those forecasters who fill in a low number of intervals in the questionnaire, thereby employing several different approaches across the panel of respondents.[14]

2.2 The choice of theoretical density function: the skew-normal family of distributions

Most of the recent literature has focused on extracting information about the first two moments of inflation expectations and neglected the importance of accounting for higher-order moments in their estimation. Allowing for potential asymmetries in the SPF distributions is however important for several reasons. First, it is likely to enhance the reliability of our estimates of the first two moments. Second, it provides a substantially larger set of information compared to previous analyses of the SPF data. The degree of asymmetry surrounding the central expectation (the so-called "balance of risks") has also generated substantial interest in recent years and, for example, plays a prominent role in the inflation density forecast published by some important central banks in the form of the so-called "fan-charts".

[13] Instead, when the presence of outliers is suspected, a negative τ is advisable to dampen their impact on the criterion (Lindsay [1994]).

[14] For instance, Engelberg et al. [2006] employ a triangular distribution for those histograms with less than three active bins and a beta for the rest. Rich and Tracy [2006] propose using information from the point estimates reported by the forecasters in the framework advocated by Wallis [2005], while D'Amico and Orphanides [2006] proposes direct fitting of the cumulative distribution of inflation uncertainty measures.

The goal of extracting as much information as possible from the reported distributions should however be weighted against the constraint imposed by the limited amount of information embodied in the reported histograms. The robustness of our proposed fitting criterion however allows us to comfortably handle even those survey replies with probability concentrated in very few bins. Nevertheless, the need to keep the exercise feasible leads us to consider a three-parameter density family and focus attention on the first three moments of the distribution.

The second novelty of this paper is the choice of a potentially skewed density function: the skew-normal distribution (see Azzalini [1985]). This family of distributions has several advantages for our purpose here. It is a relatively parsimonious density function (fully defined by three parameters). Nevertheless, under appropriate parametrisation, it provides a direct one-to-one mapping between its three parameters and the mean, variance and skewness of the distribution, which is very convenient for our exercise.[15] We therefore "let the data speak" about potential asymmetries present in the SPF data without restricting our estimates of the other moments of the distribution. Moreover, the skew-normal class, though very flexible in its shape, always remains unimodal, which is a reasonable premise when dealing with macroeconomic expectations.

More formally, the class is built, as in the case of normal distributions, by shifting and re-scaling a standard distribution with density function defined as

$$f_\lambda(z) := 2\varphi(z)\Phi(\lambda z) \qquad z \in \mathbb{R}$$

where φ and Φ are the standard normal density and distribution functions, respectively, and $\lambda \in \mathbb{R}$ is the *shape parameter*.[16] In what follows, we shall use the abbreviation $SN(\lambda)$ for the standard skew-normal distribution just defined. For practical applications, a general random variable Y is said to be skew-normal distributed when it can be written as

$$Y = \mu + \sigma \left(\frac{Z - E[Z]}{\sqrt{V(Z)}} \right) \qquad Z \sim \mathcal{S}N(\lambda)$$

The first three central moments of Y are then expressed as[17]

$$\begin{aligned} E[Y] &= \mu \\ V(Y) &= \sigma^2 \\ SK(Y) &= \gamma = \left(2b^2 - 1\right) b\delta^3 / \left(1 - b^2\delta^2\right)^{3/2} \\ &\quad \text{where} \quad b = \sqrt{2/\pi} \quad \text{and} \quad \delta = \lambda / \sqrt{\left(1 + \lambda^2\right)}. \end{aligned}$$

Figure 1 presents four density functions of skewnormal distributions of zero mean, unit variance and four different values of the shape parameter λ.[18]

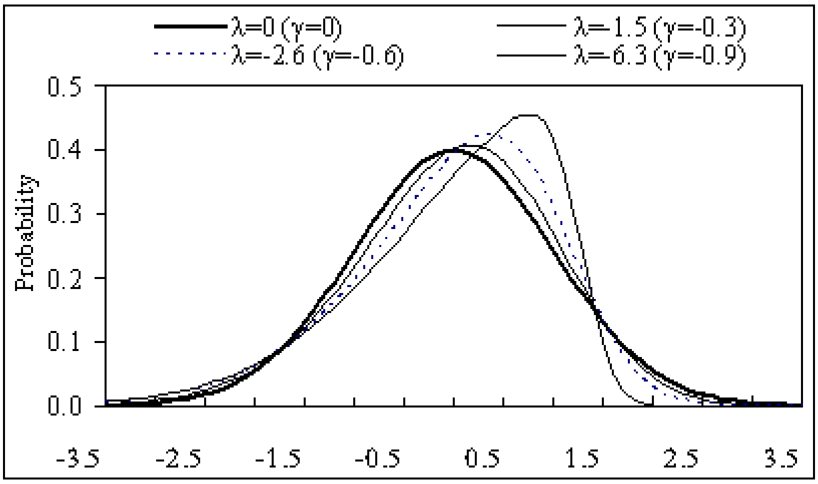
[15] We also experimented with other more parsimonious but also potentially asymmetric distributions, for example the beta and related family of distributions. However, the fact that only two parameters determine the three moments of interest imposes undesired constraints on their estimation.

[16] It is immediate to verify that for $\lambda = 0$, the skew-normal becomes the standard normal distribution. Moreover, if Z is a $\mathcal{S}N(\lambda)$-distributed random variable, then $-Z$ is a $\mathcal{S}N(-\lambda)$ random variable. The sign of λ thus simply indicates the direction of the skewness (see Azzalini [1985] for the properties of the distribution).

[17] This parameterisation avoids inference problems stemming from singularities in the Fisher information matrix (see Azzalini [1985] and Pewsey [2000]). It is interesting to note that $\gamma \in (-0.995, 0.995)$, and it is a monotonous function of the shape parameter λ, sharing its sign.

[18] The theoretical probabilities, $p_i(\mu, \sigma, \gamma) = F_Y(\alpha_{i+1}) - F_Y(\alpha_i)$, can be computed numerically us-

Figure 1 Examples of skewnormal densities



3 An assessment of our proposed methodology

This section highlights the advantages of the methodology that we described in the previous sections to estimate the key moments of the SPF probabilistic forecasts. Two pieces of information are provided. First we illustrate through qualitative examples how allowing for a potentially skewed distribution is crucial to correctly pin down the two lower-order moments of the distribution. Then we provide some quantitative evidence on those improvements.

3.1 Fitting continuos densities to the SPF histograms: some practical considerations

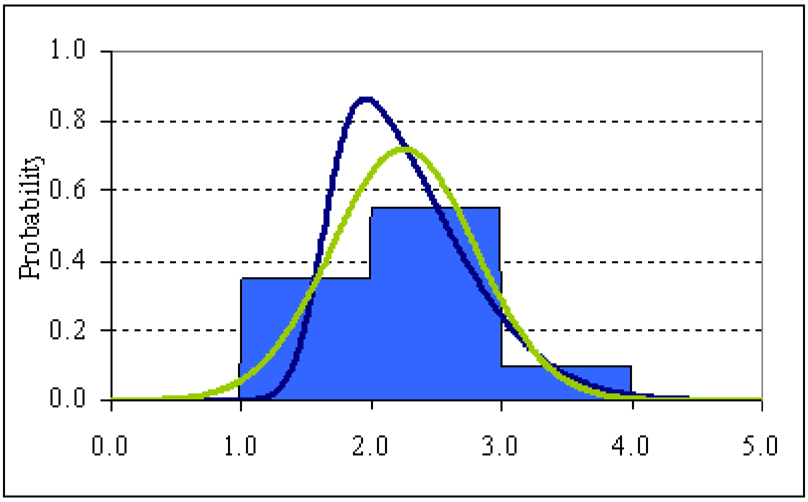
Previous approaches to fit continuos densities to the SPF histograms have encountered two related problems. First, the fitting of a continuos density to the individual responses is often complicated by the fact that some respondents only assign probabilities to a limited number of bins, possibly due to some rounding, which makes more difficult the identification of the parameters of the underlying distribution. Second, a visual inspection of the SPF histograms also suggests that a non-negligible proportion of respondents appears to report an asymmetric distribution of risks in their density forecasts. In those cases, it is immediate that imposing by assumption symmetry in the theoretical density, for example by using a two-parameter distribution such as the normal, would potentially bias the estimates of the first two moments of the distribution thereby recovered.

We have argued that the methodology proposed in this paper can overcome those two problems. Specifically the robustness to inliers of our estimator allows for taking into account the information conveyed by the empty bins surrounding those to which some positive probability has been assigned, thereby helping in the identification of the optimal shape of the distribution. This possibility is fully exploited by means of the skew-normal

ing the fact (see Azzalini[1985]) that $F_{Z,\lambda}(z) = \Phi(z) - 2T(z,\lambda)$, whereby $T(z,\lambda)$ is the Owen function, for which efficient and precise approximations exist (Patefield and Tandy[2000]), since $F_Y(\alpha) = F_{Z,\lambda}\left(\left(1-(b\delta)^2\right)^{1/2}\left(\frac{\alpha-\mu}{\sigma}\right)+b\delta\right)$.

distribution, whose flexibility allows for handling those histograms with rather skewed distribution of probabilities across their bins. Figure 2 illustrates with an actual example.

Figure 2: Skew-normal versus normal fitting



The inflation forecast corresponds to forecaster 541 in the 2005Q3 survey round.

It is immediate from the picture that the shape of the histogram, although relatively bell-shaped, can hardly be reconciliated with a symmetric distribution. The skew-normal matches the features of the histogram better . That particular example also shows that the increased flexibility stemming from the endogenous skew of our theoretical distribution is very important for extracting reliable estimates of the first two moments of the probabilistic forecast. In particular, in the presence of asymmetry, the difference in the estimation of the variance of the distribution is already visible in Figure 2. The next section provides some additional quantitative evidence.

3.2 Accuracy gains of the proposed methodology: Monte Carlo evidence

The purpose of our Monte Carlo experiments is twofold. First, to guide our selection of estimator within the family of power divergence estimators ($PDE(\tau^*)$ henceforth), i.e. the appropriate choice of τ for the SPF-type of data. Second, to provide some quantitative evidence on the relative improvements from the methodological contributions proposed in this paper (namely the fitting criterion $PDE(\tau^*)$ and the skew-normal distribution) compared with the (unweighted) LS criterion and the normal distribution employed in earlier literature.

The Monte Carlo experiments are designed as follows. Pseudo-random draws are taken from different underlying densities that range from a skewnormal with a high degree of asymmetry ($\gamma = 0.9$) to the symmetric normal distribution ($\gamma = 0$), and histograms are formed on the basis of the relative frequencies over a grid of intervals similar to the actual SPF questionnaire.[19] The parameters of the underlying distributions are then estimated from those histograms using four different combinations of theoretical densities and fitting

[19] Pseudo-random draws from the SN distribution were obtained following Henze [1986]. The range of intervals is $(-\infty, 0)$,$[0, 1)$,$[1, 2) \ldots$,$[7, 8)$,$[8, +\infty)$.

criteria (Normal and Skew-normal densities, and LS and $PDE(\tau^*)$). Therefore, these numerical results allow for assessing the accuracy gains stemming from the separate or combined use of the two methodological contributions introduced in this paper.

First we focus on the qualitative implications of the choice of τ by examining estimation errors. Ideally, the chosen parameter should simultaneously minimize estimation errors incurred for all three density parameters (μ, σ, γ), but in practice such optimality is unlikely. Our simulations suggest that the performance of the estimators is relatively similar within the $(-0.5, 1)$ range for the three estimated parameters.[20] Since a low but positive value of τ seems reasonable to cope with the inlier-related problems discussed in the previous section, we chose $\tau^* = 0.2$. Our proposed approach is therefore fully specified as *PDE*(τ=0.2, *SN*)). We now turn to analyse its performance in detail.

The results of the Monte Carlo simulations show significant accuracy gains from each of the two methodological contributions introduced in previous sections, and even stronger ones when both of them are used jointly. Such gains are illustrated in the four charts below. We here focus on the results based on 50-draw samples and a few key results, but a comprehensive sensitivity analysis of the accuracy gains in terms of estimation criterion, true underlying distribution and sample size can be found in the Appendix.[21]

First, the estimator $PDE(\tau^*)$ proposed in this paper clearly outperforms (unweighted) LS for both the mean (μ) and standard deviation (σ), thereby providing some quantitative support for our theoretical argumentation that both kinds of estimators are consistent but only minimum disparity estimators are efficient. Figure 3 illustrates the gains in terms of the percentage decline in the mean square error (MSE) from using the $PDE(\tau^*)$ criterion with respect to LS when fitting a *SN* density to samples generated from distributions with different degrees of asymmetry. Both for the mean and for the standard deviation the decline in MSE is of at least 20% in all cases. Moreover, these gains appear to be robust not only with respect to the degree of asymmetry present in the data but also to the choice of model distribution to fit the data: while the results depicted in Figure 4 are based on the use of the *SN* as model distribution, the comparison of the results of using the normal distribution with both fitting criteria reveal similar accuracy gains (see Appendix).[22]

[20] Alternative values of the mean and the standard deviation were explored and led to the same qualitative conclusions. Besides, different sample sizes were also considered, but apart from the expected convergence at the usual n^{-1} rate, no noticeable differences were found. For further information see Annex III.

[21] Specifically, each of the columns of the tables in the annex respectively correspond to the power distance estimator (PDE) using a skew-normal density (PDE(τ^*), SN)), i.e. our proposed methodology, the power distance estimator using a normal density (PDE(τ^*), N), the standard least squares criterion using the Skew-normal density (LS (SN)) and the least squares criterion using the normal density (LS (N)). Their performance is measured by mean square errors (MSE), mean absolute errors (MAE), and the empirical bias obtained from 1000 simulated samples of the different sizes.

[22] Compare columns 1 and 3 (for mean) and columns 5 and 7 (for variance).

Figure 3: Fitting criterion gains (MSE)

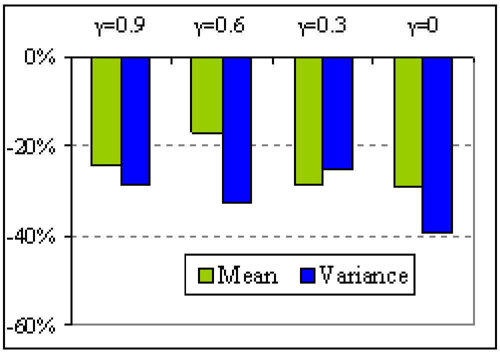


Figure 4: Density gains (MSE)

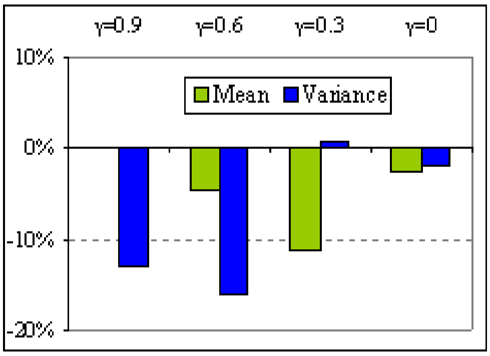


Figure 5: Overall gains from LS,N (MSE)

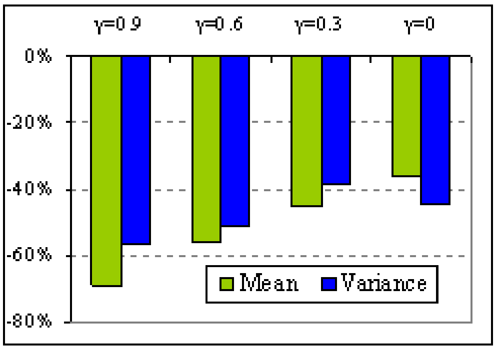
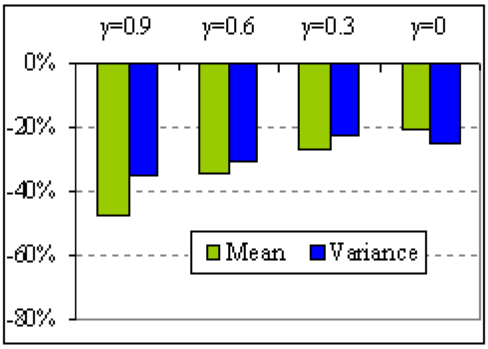


Figure 6: Overall gains from LS,N (MAE)



Note: The charts depict the percentage reduction in the estimation error (mean absolute error (MAE) and mean square error (MSE)) of our methodology with respect to the least squares (LS) fitting criterion and/or the standard normal (N) density for the mean and the variance of the distribution. Specifically, Figure 3 reports the percentage reduction in the MSE from using our fitting criterion instead of LS for the Skew-Normal (SN) distribution, Figure 4 for using the SN instead of the N with our fitting criterion, and Figures 5 and 6 from using our fitting criterion and the SN instead of LS and the N distribtuion. The results are based on 1000 simulations from skew-normal distributions with four different degrees of skewness, as indicated in the horizontal axis.

Second, there are strong accuracy gains in the estimation of the key moments of the distribution stemming from the use of a potentially asymmetric theoretical density function – such as the *SN* proposed here – rather than imposing symmetry in the estimation by the use of the normal density. Figure 4 depicts the percentage gains in terms of MSE from using the *SN* rather than the normal distribution to fit the data. These gains seem to be more limited, but is has also to be taken into account that the reported figures are based on the use of our proposed fitting criterion. Again, if a less efficient method, such as *LS*, were to be used with both distributions the gains from employing the *SN* would be significantly larger (see Appendix).[23] It is particularly noteworthy that the SN still performs well in the estimation of the two first moments even in the case of symmetry in the true DGP. This is very encouraging, for it is difficult to assess *a priori* the degree of asymmetry present in the actual SPF data, and it could be possible that, since the *SN* model is overparameterised in the case of fitting symmetric data, estimation errors would be higher than using the symmetric normal distribution. It however appears that the flexibility of the SN clearly overcomes potential problems related to overparameterisation. Particularly in small samples (i.e. discretisations based on relatively small number of draws) it has to be taken into account that pinning down the true parameters of the DGP is problematic, and our results suggest that the SN is capable of accommodating deviations

[23] Compare columns 3 and 4 (for mean) and columns 7 and 8 (for variance).

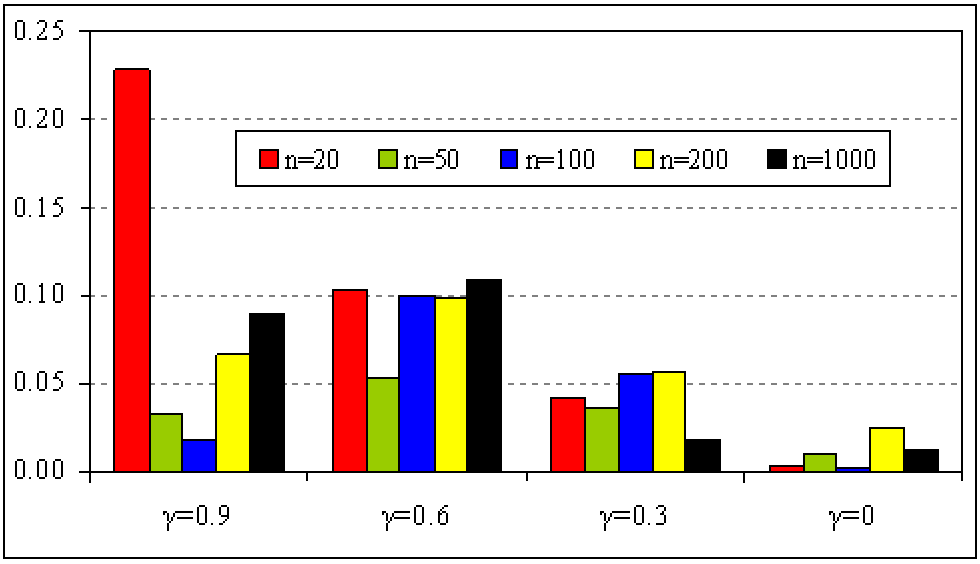
from normality through its shape parameter without compromising the estimation of the first two moments.

Third, the joint use of the two methodological contributions discussed in previous sections, namely $PDE(\tau^*)$ as fitting criterion and SN as underlying distribution, yields very strong performance gains over LS fitting assuming a normal distribution. Figures 5 and 6 illustrate the accuracy gains in terms of mean square errors (MSE) and mean absolute errors (MAE) respectively. Both for the mean and the standard deviation the improvement in terms of the MSE is of about 40%. In terms of the MAE, the improvement, although somewhat lower than for the MSE, is also substantial (above 20% in all cases).

In sum, the results of our Monte Carlo exercise confirm that the methodology proposed in this paper provides quite accurate estimates for the first two moments of the SPF distributions. However, it is also important to note that our use of the SN as theoretical distribution also provides information about the third moment of the distribution. As pointed out before, most of the existing literature has restricted attention exclusively to the first two moments of inflation expectations. Therefore, as this is a novelty of our analysis, we also assess the accuracy of the proposed methodology to pin down the skewness of the distribution.

Figure 7 depicts the mean estimation error for different sample sizes. The data generating process is always a SN, where we try different values of the true γ parameter in order to gauge the performance of our method under different degrees of asymmetry in the data.[24]

Figure 7: Estimation of skewness parameter: forecast error (absolute values)



The reported mean absolute errors are based on 1000 simulations for each distribution.

Several features are noticeable from the Figure. First, estimation errors are rather small in all cases: they are below 0.1 on average, the only exception being the combination of rather small sample size (n=20) and highly skewed distribution (γ=0.9), a very challenging combination that should nonetheless be consider a rather extreme hypothetical

[24]Note that it is sufficient to conduct the simulations for zero and positive values of gamma due the symmetry of the skewnormal distribution with respect to γ.

case. Second, in general, it appears to be more challenging to pin down the skewness for the case of highly asymmetrical distributions. For instance, estimation errors are negligible for symmetric distributions even for rather small samples sizes, but appear to increase with the degree of asymmetry present in the true data independently of the sample size. Finally, a tendency towards a slight positive bias is also observable in our simulations. This tendency appears to be related to the parametrisation discussed in Section 2.2. This is particularly relevant in the case of heavily skewed distributions (γ=0.9), in which numerical estimates tend to be pushed towards the upper bound (the lower bound for negative γ). Therefore, this is not likely to affect in any strong way the qualitative information content of our estimates, i.e. the presence of marked skewness in inflation expectations would be identified in our approach.

Overall, the first two moments are easier to pin down than the skewness parameter, but our simulations suggest that the asymptotic rate of convergence for the estimated skewness parameter is linear, despite its (finite sample) variance being higher than for the first two moments, in line with existing results (Pewsey [2000]).

4 Risks in inflation expectations

This section investigates the information content of the actual SPF probability distributions in the period 1968Q4-2006Q1. The SPF questionnaire requests density forecasts for inflation over the current (since 1981Q3 also the next) calendar year for four consecutive quarters. Using our methodology we estimate the forecasters's baseline scenario (for example the mean of the distribution), as well as the uncertainty surrounding it (i.e. the variance) and the risk assessment (i.e. skewness) associated with that baseline scenario from the reported probability distributions. Following the notation introduced in section 2.1 for a given survey, we apply the methodology described above both to each individual response, $\{p_{ik}\}_{i=0,...,I,k=1,...,K}$ for all K forecasters and to the *consensus* (aggregate) distribution calculated as a simple average of responses across forecasters, $p_i = \frac{1}{K}\sum_{k=1}^{K}(p_{ik})$.

Regardless of whether one focuses on the *consensus* or the individual distributions, the interpretation of the changes in the distribution of inflation expectations over time is complicated by the fact that the SPF forecast horizon decreases over time as the survey takes place closer to the end of the year for which the forecast is requested.[25] Time-series evidence is however interesting to assess how the expected inflation rate, and the uncertainty and risks associated to it, have co-moved over time. To that end, we present evidence of those comovements based on expectations over a fixed forecast horizon by using the results for the surveys taking place in the first quarter of the year, i.e. reporting forecast over a horizon of about four quarters ahead. Obviously this evidence has the shortcoming of being based on annual frequency data. To shed additional light on the robustness of those results, an alternative is to combine forecasts for the current and the next-year horizon after the introduction of the latter in the questionnaire in 1981. Specifically, an about *four-quarter-horizon forecast* at quarterly frequency can be constructed by combining the

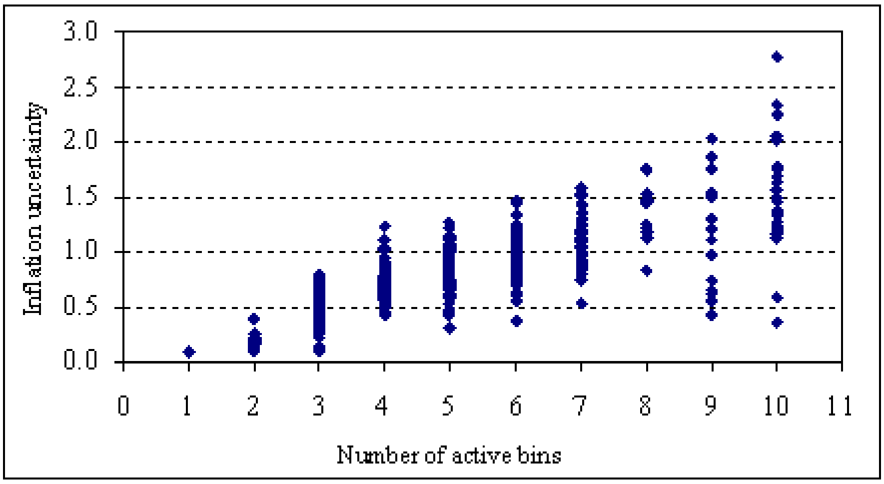
[25] There were also some occasional mistakes in the horizons requested in the questionnaire but they do not affect the results presented in this paper. For example, in certain surveys mostly prior to 1981Q3, most often in the fourth quarter, the probability variables referred to the percent change in the GNP deflator in the following year, rather than the current year. The surveys in which these happens are 1968Q4, 1969Q4, 1970Q4, 1971Q4, 1972Q3 & Q4, 1973Q4, 1975Q4, 1976Q4, 1977Q4, 1978Q4, and 1979Q2, Q3 and Q4. The opposite situation appeared to occur in the 1985Q1 and 1986Q1 surveys. See http://www.philadelphiafed.org/econ/spf/index.html for details.

current year forecast in first quarter (i.e. about four-quarters ahead) and in the second quarter (i.e. about three quarters ahead) with the next year forecasts in the third quarter (i.e. about six quarters ahead) and in the four quarter (i.e. about five quarters ahead).

4.1 Measuring inflation uncertainty

The analysis of the uncertainty surrounding the baseline scenario of a forecast has gained substantial attention in recent years. In this regard, the probability distributions of the SPF are the best available source of information about the existing degree of uncertainty of inflation expectations, because uncertainty, in a strict sense, is directly reflected in the probabilities that a forecaster assigns to the possible alternative values of the predicted variable: the tighter the distribution the lower the uncertainty.

Figure 8: Estimated uncertainty and number of active intervals reported



Results from individual probability distributions 1992Q1-2006Q1

Figure 8 depicts the relationship between our estimates of individual uncertainty (as measured by the variance of the estimated distribution) against the number of intervals reported by the survey participants in the period 1992Q1-2006Q1, the last period over which interval definition and width has been stable. Consistent with common intuition there is a clear positive relationship between both series. However it is also clear that, despite assigning some positive probability to the same number of intervals, forecasters significantly disagree on the distribution of probability mass assigned to those intervals, thereby leading to quite strong spread in our measures of uncertainty.

However, while for an individual forecast it is straightforward to measure uncertainty, in the case of the SPF, once all individual forecasts are taken into account it is not clear what the optimal measure of uncertainty is and this is most likely to depend on the specific purpose of the exercise.[26] For example, given that many surveys only provide a single point prediction, there is a long stream of literature that has focused on the information content of the dispersion of those point estimates –the so-called *disagreement*– as alternative indicator.[27] The SPF information however allows for an analysis of those two alternative measures of uncertainty.

[26] See Giordani and Söderlind[2003] for a recent detailed discussion.
[27] For a critical assessment see Rich and Butler [1998].

The second moment decomposition (ANOVA) has frequently been read as explaining the variance of the aggregated distribution as a sum of the average uncertainty of individual expectations and disagreement among them. Formally:

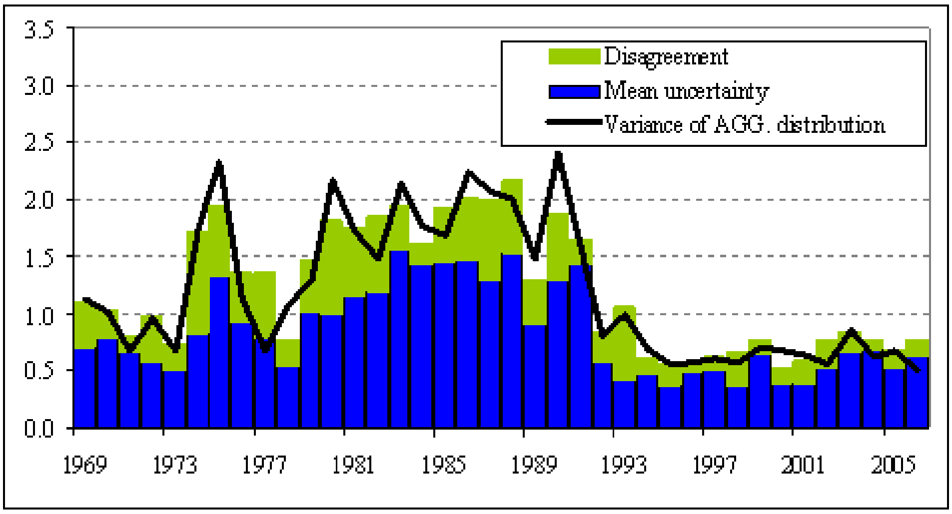
$$Var(\pi) = E\left[(\pi - E[\pi])^2\right] = E\left[Var\left(\pi|F\right)\right] + Var\left(E[\pi|F]\right)$$

where, as in section 2.1, F is a discrete random variable that selects a forecaster F_k among the panel with a certain probability. If F assigns equal probability to each forecaster, i.e. if all forecasts are equally weighted

$$Var(\pi) = \frac{1}{K}\sum_{k=1}^{K} Var_i(\pi) + \frac{1}{K}\sum_{k=1}^{K}(E[\pi]_i - E[\pi])^2 \tag{4}$$

Equation (4) holds in population terms independently of distributional assumptions. Empirically, one can estimate the left-hand side from the aggregated survey response as the second moment of the estimated density. The right-hand side, in turn, can be derived from the moments of the estimated individual density functions.

Figure 9: Variance of the aggregate distribution, mean uncertainty and disagreement



Decomposition of the variance of the aggregate distribution for the current year forecast (Q1).

Figure 9 shows that inflation uncertainty has exhibited significant fluctuations over the last 40 years. The sharp increases following the two oil price crises in the 1970s are directly identifiable, but some other spikes in the 1980s are also clearly visible in the graph. In addition, the strong decline in inflation uncertainty observed since the early 1990s is also noticeable, and could be one of the factors that help explain the lower premia, in particular inflation risk premia, allegedly embodied in bond yields in recent years. Although such moderation is likely to reflect a number of factors, the enhanced credibility of monetary policy appears to be one of them. We will return to this interpretation in the next section.

Additional insights when considering the appropriate measure of uncertainty about inflation expectations can also be drawn from the decomposition of inflation uncertainty depicted in Figure 9. First, the decomposition of the variance of the aggregate distribution shows that mean uncertainty, a more theoretically-appealing measure of the perceived inflation uncertainty, is clearly the main component of the degree of uncertainty reflected in the consensus probability distribution. Disagreement among individuals with respect to the mean forecast instead appears to play a secondary role. Indeed, the contribution of disagreement to the variance of the consensus probability distribution appears to have decreased over time and seems to be relatively minor since the early 1990s.

Second, the three measures depicted in Figure 9 have experienced a significant decline since early 1990s notwithstanding that moderation in uncertainty measures. Again it seems apparent that most of the decline in aggregate uncertainty stems from the decline in mean uncertainty. Confirming that visual impression, simple measures of comovement between the three measures of uncertainty show that the correlation between the variance of the aggregate distribution and mean uncertainty extracted from the individual distributions is higher than that with disagreement (see Table 1), a result that appears robust to data frequency and period of study.

Table 1 : Correlation among alternative measures of inflation uncertainty

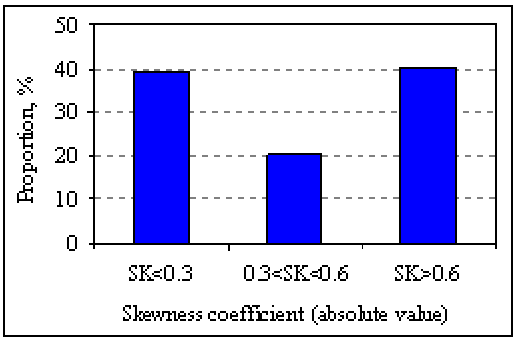
Q1surveys (1969Q1-2006Q1, 37obs)	Variance of agg. distribution	Mean uncertainty	Disagreement
Variance of aggregate distribution	1.00		
Mean uncertainty	0.90	1.00	
Disagreement	0.74	0.54	1.00
Quarterly series (1981Q1-2006Q1, 101obs)			
Variance of aggregate distribution	1.00		
Mean uncertainty	0.89	1.00	
Disagreement	0.79	0.64	1.00

4.2 Assessing asymmetries in inflation risks

While the uncertainty surrounding inflation expectations is clearly informative, a reliable assessment of how the risks surrounding the baseline scenario are perceived is also of significant relevance for policymakers and market participants alike. By overcoming the symmetry assumption, our skewnormal model allows to gain additional insights on the risk assessment embodied in the inflation forecasts.

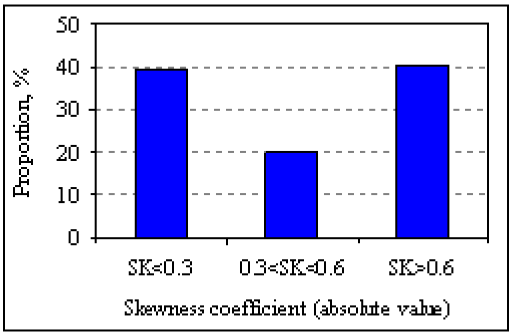
The reported density forecasts present an important degree of asymmetry (see Figures 10 and 11). Indeed, although this information has been often neglected so far, one of the most striking features of the SPF probability distributions appears to be their asymmetry. Our estimates of the Pearson skewness parameter γ suggest that not only more than 60% of the individual probability distributions reported over the whole sample 1968Q4-2006Q1 show asymmetries higher than 0.3 (in absolute value), but also more than 60% of the aggregate probability distributions constructed out of them do so.

Figure 10: Skewness (aggregate PDFs)



Based on 247 estimated aggregate densities, 1968Q4-2006Q1.

Figure 11: Skewness (individual PDFs)



Based on 8204 estimated individual densities, 1968Q4-2006Q1.

This evidence is particularly important for two reasons. First, as shown in Section 3.2, allowing for asymmetry in the fitted probability distribution is not only worthwhile in its own right but also improves the accuracy in the estimation of the first two moments of the distributions. Indeed, these results provide strong support for our choice of a potentially skewed distribution and the reliability of the estimates of the first two moments of the distributions characterising inflation expectations presented here. Second, accounting for the fact that a very large proportion of individual and aggregate probability distributions show significant asymmetry raises a number of new interesting questions on the information content of the SPF inflation expectations. We analyse some of them below.[28]

Inflation expectations are often reported by a single number, a *point estimate.* Indeed, with the notable exception of the SPF macroeconomic surveys tend to request participants to provide just a single number, and a "consensus point estimate" is often calculated as the average (or median) of the point estimates reported by survey participants in the same way as the results of the aggregate (across forecasters) probability distribution are taken as representing the consensus density forecast. However, the presence of asymmetry in the probability distributions raises the issue of which location measure is more relevant as single point estimate. The larger amount of information extracted from the methodology presented in this paper is crucial to gain some additional insights in this matter. In terms of the individual probability distribution the proper location measure to report is directly related to the forecasters' loss functions. On the one hand, for what may be considered standard (quadratic) loss functions the *mean* would be the most relevant location measure, as would be the median for an L_1 loss function.[29] On the other hand, forecasters might report what they consider the most likely value i.e. the *mode* rather than the *mean* of the distribution. In any case, since it does not seem possible to gather evidence on the true forecasters' loss functions, here we use our methodology to shed some light on this issue by comparing our estimates of the individual mean and mode.[30]

We investigated which of the two location measures varies more with the arrival of new information and the revision of the forecast for a given horizon over a forecasting cycle (i.e. the series of forecasts reported for a given calendar year in several survey rounds).

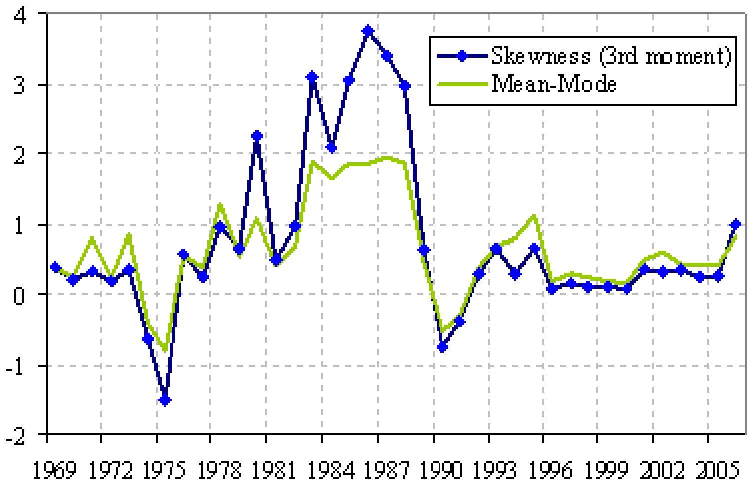
[28] For a more detailed analysis of the degree of heterogeneity in inflation expectations and its implication for the interpretation of survey results see García and Manzanares ([?])

[29] A quadratic loss function would penalise (in proportion to) the square of the forecast error, while a L_1 loss funtion penalises (in proportion to) the absolute value of the forecast error.

[30] A detailed comparison of the information content of the point estimates and the density-forecast based central tendency measures derived using othe methodology presented here can be found in the companion paper Garcia and Manzanares [2007b].

The standard deviations of means and modes of the probability distributions show that both change over each forecasting cycle as we get closer to the forecast horizon and new information arrives. However, the standard deviation of the modes remains systematically higher than that of the means over the whole sample period, which may suggest that the means should be interpreted as the forecasters' baseline scenario while the mode, as most likely value of the density forecast, may be more influenced by the re-assessment of the balance of risks associated to the forecast. Some additional evidence tends to support that conclusion. Specifically, the difference between the mean and the mode of the distributions, a simple measure of asymmetry in the expectations, co-moves strongly with the more formal skewness measure based on the third moment of the estimated probability distribution. Figure 12 illustrates for the case of the aggregate probability distributions in Q1 surveys.[31]

Figure 12: Alternative measures of skew (aggregate distribution)



The perceived balance of risks surrounding inflation forecasts also seems to have changed significantly over time. Such fluctuations appear to be more difficult to interpret and identify with specific historical episodes than the ones of inflation uncertainty reported above though. In particular, the perception of significant upside risks in the mid-1980s stands out as the main feature of those fluctuations. We come back below to the significance of that phenomenon with the help of additional historical evidence, and focus now on its decomposition.

Our analysis also allows for an investigation of how the balance of risks in the aggregate distribution may emerge. By making use of the moments of the individual density forecasts

[31] SPF panelists also provide forecasts for the price level index over the next six quarters, which allow for a calculation of inflation point estimates for the same horizon of the current-year probability distribution. Although their average is often assumed to be the mean of the aggregate distribution, this would only be true the case *if* the point estimates were the mean of the individual distribution. Our results (not reported here) provide additional support to recent evidence by Engelberg *et al.*[2006] suggesting that the individual forecasters' point estimates do not always coincide with our estimates of the mean (nor the mode) of their probability distributions, thereby casting doubts on the validity of the average of the point estimates as estimate for the mean of the aggregate distribution.

in a similar way as for the variance, the third central moment of the aggregate inflation expectations (denoted $SK(\pi)$) can be decomposed into three components as follows[32]

$$SK(\pi)=E\left[(\pi-E[\pi])^3\right]=E\left[SK\left(\pi|F\right)\right]+SK\left(E[\pi|F]\right)+3\ E\left(Var\left(\pi|F\right)\left(E[\pi|F]-E[\pi]\right)\right)$$

The first and second components, in analogous fashion to the well-known variance decomposition, correspond to the average individual skewness and to the skewness of the distribution of individual means. The skew decomposition however requires of a third additional component: the covariance between the distance of the individual forecaster's mean to the consensus mean in that survey round and the variance of that forecaster. So, intuitively, it reflects the impact that the different degrees of uncertainty among forecasters can have if it is correlated with their disagreement with respect to the consensus forecast. For instance, this third term would vanish if all forecasters had the same degree of uncertainty.

The decomposition of the degree of skewness in the aggregate distribution makes evident that the mapping between aggregate and individual skewness is far from simple though. On the contrary, it highlights the importance of understanding the cross-sectional (across forecasters) distributions of the key moments of the individual probability distributions so as to understand the consensus one. For instance, the comovement among its three components also suggests that inferring the sources of the balance of risk embodied in the aggregate distribution is rather challenging and the importance of the three theoretical components is likely to change over time (see Table 2).

Table 2 : Comovement among measures of skewness in inflation expectations

Q1surveys (37obs)	Aggregate skewness	Mean skewness	Skewness of means	Cov. means and variances
Aggregate skewness	1.00			
Mean skewness	0.28	1.00		
Skewness of means	0.55	0.03	1.00	
Cov. means and variances	0.78	0.39	0.24	1.00

4.3 Inflation risks and the central tendency of inflation expectations

In the previous sections we have focused on the uncertainty and asymmetry of the SPF inflation forecasts and described in detail some of their key features focusing on the relationship between the individual density forecasts provided by SPF panelists and the one resulting from averaging them into a combined (or aggregate) density forecast. In this section we provide some formal evidence of the link between mean inflation expectations and the uncertainty and risk assessment measures derived in this paper. We focus on our estimates of the first three moments of the aggregate density forecasts, and seek for stylised facts about how perceived inflation risks changed in relation to mean inflation expectations, thereby providing an assessment of the evolution of density forecasts of inflation over the last four decades.

[32] For the sake of exposition, in this subsection we will use the term *skewness* to denote the non-normalised central third moment, instead of the Pearson's skewness coefficient used in the previous sections.

Figure 13 depicts mean inflation expectations together with the inflation uncertainty and the degree of skewness associated to those expectations, thereby providing some preliminary evidence of the relationship between them. Inflation uncertainty appears positively related to the level of expected inflation, while a weak negative relationship appears to exist between perceived skewness and inflation.

Figure 13: Expected inflation and inflation risks

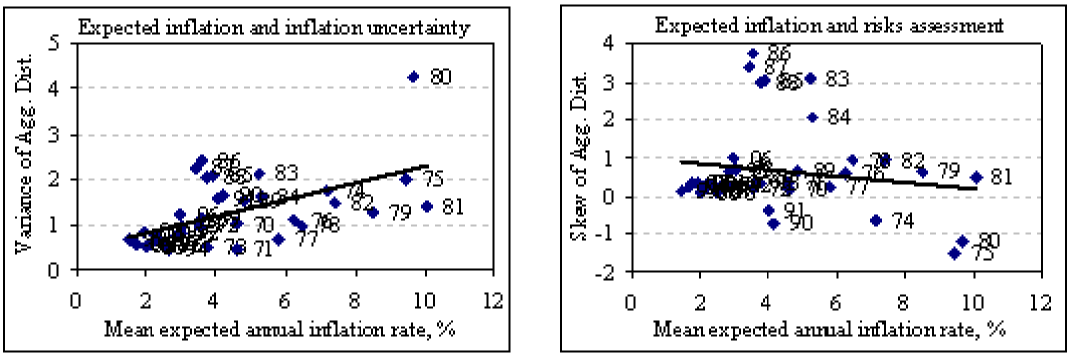


Table 3 below assess the comovements among the first three moments of inflation expectations.[33] Several results are worth noting. First, a comparison of the variances of the first three moments of inflation expectations (reported along the principal diagonal in Table 3) suggests that central tendency measures have been the most volatile over the sample as whole (results for other central tendency measures, as the average or the median of individual means, are qualitatively equivalent). As regards higher-order moments, although both uncertainty and skewness showed significant increases in the 1980s, the level of uncertainty surrounding inflation expectations appears to have been much less volatile than the asymmetry in the perceived inflation risks. Second, in terms of comovement between moments, mean inflation expectations seem to be fairly strongly correlated with uncertainty measures, in particular with disagreement, but little with the measures of asymmetry in risks. However, possibly reflecting the upward movement both higher-order movements experienced in the 1980s and the decline thereafter,inflation uncertainty appears to be fairly strongly and positively correlated with skewness in inflation expectations.

Table 3: Correlation between the first three moments of inflation expectations

	Agg. Mean	Agg. Variance	Mean uncertainty	Disagreement	Agg. Skew	Mean-Mode
Agg. Mean	(5.45)					
Agg. Variance .	0.59	(0.37)				
Mean uncertainty	0.48	0.88	(0.15)			
Disagreement	0.64	0.73	0.53	(0.05)		
Agg. Skew	0.07	0.54	0.60	0.31	(1.12)	
Mean-Mode	-0.02	0.42	0.53	0.21	0.98	(0.40)

Q1 surveys, 38 observations 1981Q1-2006Q1. Figures in main diagonal denote variances of the series.

[33] Results from the use of the quarterly time series combining current and next year forecasts were qualitatively similar. They are therefore omitted here but available upon request.

Table 4 reports the results of some bivariate linear regressions to assess the statistical significance of the correlations reported in the previous table. In order to provide some robustness and further insights about the relationship between the key attributes of inflation expectations we employ the quarterly time series introduced above for the sample 1981Q1-2006Q1, as well as an additional central tendency proxy, the average of individual means. The results show that the level of inflation does significantly contribute to partly explain movements in inflation uncertainty, which, together with the correlations shown above, can provide support from the existence of a positive relationship between the level of a variable and the uncertainty surrounding its forecast, a belief usually taken for granted. The results in Tables 3 and 4 show that such a relationship, although strongly significant, is however limited (the $\overline{R}^2$ in the bivariate regressions are about 0.4), suggesting that a significant part of the variation in the uncertainty surrounding inflation expectations is explained by other factors different that the level of expected inflation. Instead, the level of inflation shows very limited explanatory power for the asymmetries in the risk assessment.

As regards the relationship between inflation uncertainty and the risk assessment embodied in inflation forecasts, although all three measures of uncertainty derived here are capable of partly explaining asymmetries in perceived inflation risks, the explanatory power of the variance of the aggregate distribution, and particularly its main component mean uncertainty, seems to be stronger than that of disagreement.

Table 4: Level, uncertainty and risk assessment in inflation expectations

	Uncertainty measures			Risk assessment measures	
	Variance Agg. Dist	Mean uncertainty	Disagreement	Skew Agg. Dist.	Mean-Mode
Mean of Agg. Dist.	0.24**	0.18**	0.08**	0.10	0.03
Average of means	0.24**	0.18**	0.08**	0.10	0.03
Variance Agg. Dist.				0.85**	0.40*
Mean uncertainty				1.12**	0.54**
Disagreement				1.73*	0.82

Each cell reports a regression of the risk measures on the first column ones, using a quarterly series 1981Q1-2006Q1.
** and *denote significance at 1% and 5% level respectively (Newey-West corrected standard errors upto a year)

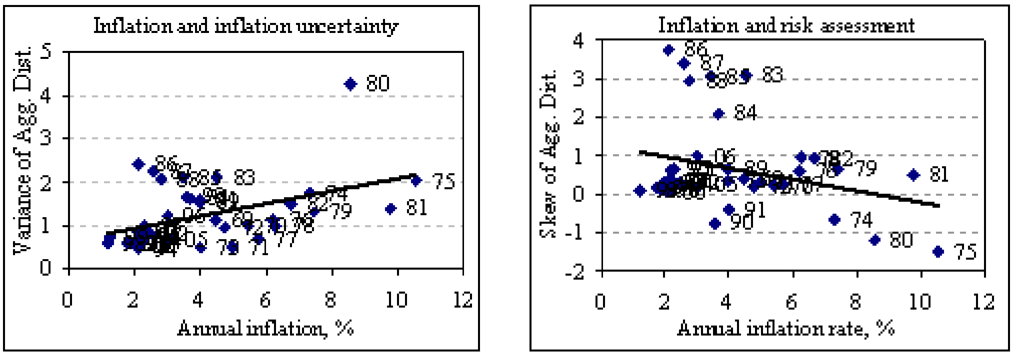
Overall, we interpret the evidence presented in this section as suggesting that although there is some degree of comovement among the three first moments of inflation expectations, limiting the assessment of inflation expectations to the mean, or any kind of central tendency measure, is not likely to be sufficient to characterise all the variation in inflation expectations.

4.4 Inflation risks and the macroeconomy

We now turn to investigate the link of the inflation risks measures derived from those combined forecasts and other key macroeconomic variables. We first investigate the role of the level of realised inflation in driving the inflation risk measures discussed above.

A positive relationship between the level of realised inflation and inflation uncertainty confirming that higher inflation has been usually associated with higher inflation uncertainty is evident in Figure 14. However, that evidence also suggests that there were some periods, mainly in the 1980s, in which, given the observed inflation rate, inflation uncertainty appeared to be "abnormally" high, either for a brief period (around 1980 following the second oil crisis) or over a more protracted period in the second half of the 1980s. This evidence deserves some close attention and we will come back to an interpretation of those episodes in the next section with the help of additional information.

Figure 14: Inflation and inflation risks



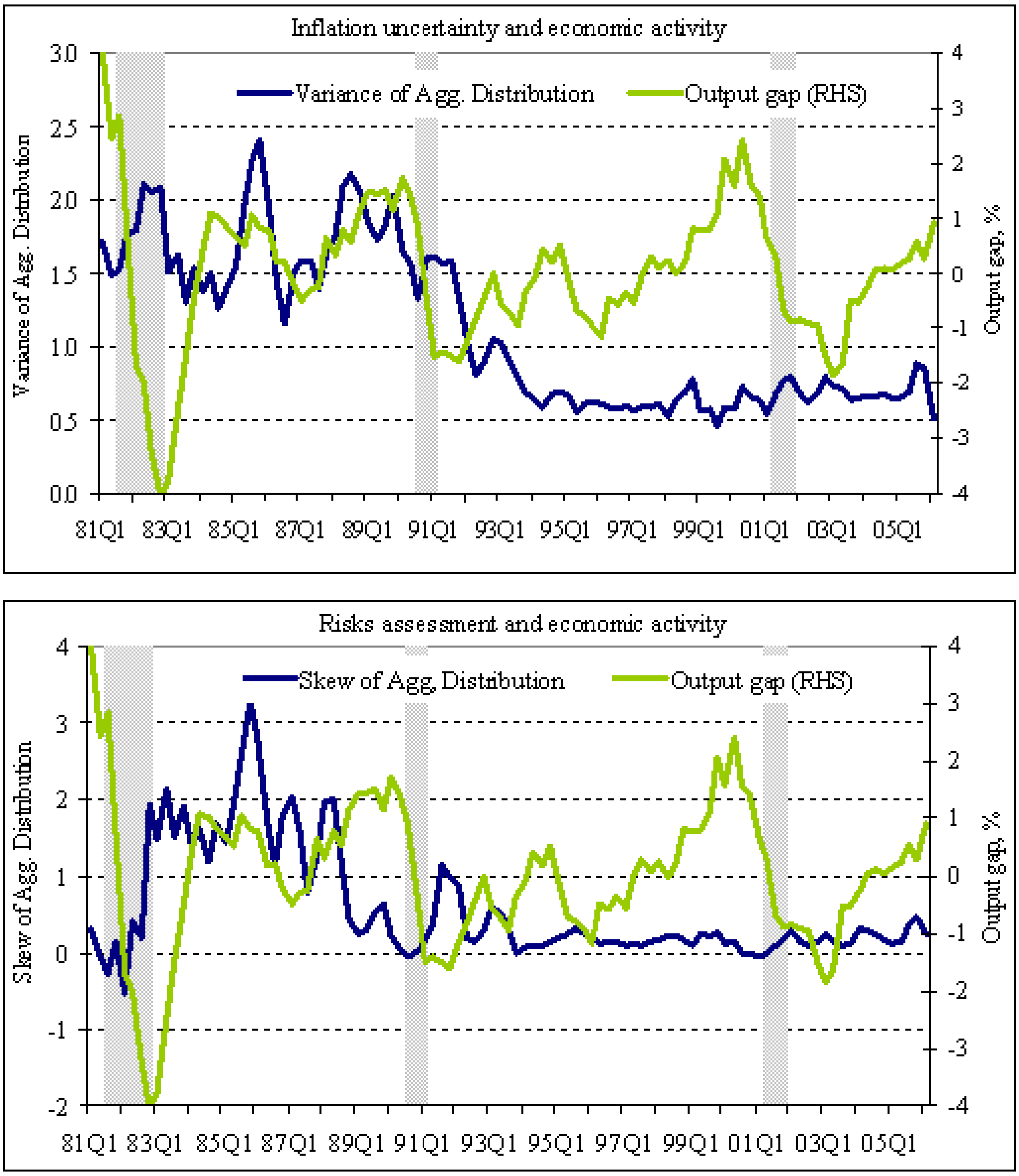
As regards forecasters' assessment of the asymmetries perceived in inflation risks, as mentioned before it is somewhat surprising that SPF density forecasts have usually exhibited positive skew and only in very few occasions over the last four decades downside risks appeared to be perceived as predominant. The negative relationship with the level of inflation apparent from Figure 14 seems to reflect in particular that a rather high inflation, and a high level of (mean) inflation expectations (see also Figure 13), as in 1974-75 and 1980, were not perceived as long-lasting, and consequently SPF panelists tended to provide negatively skewed forecasts around their baseline scenarios. However, it is noticeable that negative skew also happened at more moderate rates of inflation as in 1990-91, suggesting that negative skew in inflation forecasts may be capturing other factors beyond the actual level of actual or expected inflation. In addition, as for the case of inflation uncertainty, the period 1984-88 stands out as peculiar in terms of forecasters' risk assessment, as fairly high upside risks were associated to what appear-to-be average levels of inflation.

We next investigate how inflation risks evolve in relationship with the business cycle. Figure 15 shows two graphs depicting inflation risks together with two standard measures of excess capacity: (i) an HP-filtered output gap measure for the (log of) actual chain-weighted real output; (ii) episodes of economic contraction according to the NBER Business Cycle Dating Committee, represented as shaded areas.

The figure suggests an interesting assessment of inflation risks around recessions. First, the positive skew surrounding the reported expected inflation tends to decline ahead of the dated recessions, becoming more balanced or even turning negative, indicating that downside risks become predominant ahead of a recession. This suggests a strong business cycle response of inflation expectations, as not only point forecasts of inflation tend to decline with lower real economic activity but also the associated balance of risks tends to respond to business cycle conditions as the probability assigned to a higher-than-expected inflation outcomes is reduced. Second, in the three available episodes of contracting real GDP in the sample, inflation uncertainty also declined ahead of the recession, while at the same time there is evidence suggesting that during the recession periods uncertainty about future inflation rate tends to rise, most likely related to the uncertainty about the strength of the forthcoming recovery.

Outside recession periods there are also some interesting insights from a comparison of movements in the output gap and our measures of inflation risks. Specifically, while an output gap widening appeared to have some positive impact on both inflation uncertainty and upside risks in the 1980s, that relationship seems to have weakened significantly since the early 1990s. For instance, the robust and protracted expansion of late 1990s appeared to have no impact on perceived inflation risks.

Figure 15: Inflation risks and economic activity



Output gap is the HP-filtered. Inflation uncertainty and skew are smoothed (five-quarter centred moving averages).

Table 5 investigates more formally the relationship between our inflation risk measures and the macroeconomic situation by reporting regressions of the variance of the aggregate distribution and its two estimated components, mean uncertainty and disagreement, and the two measures of risk assessment considered before, skewness and the distance between the mean and the mode of the aggregate distribution, on several macroeconomic variables. Specifically, together with the level of inflation we also consider the (squared) change of inflation as proxy for the volatility of inflation as nominal variables, and three real economy variables: the output gap measure considered above, and the probability of recession and the (mean) uncertainty about real GDP growth (four-quarters ahead) as reported by SPF panelist. Panel A in the table report bivariate regressions for each of those variables while panels B to C control for actual inflation and (mean) expected inflation respectively to assess the robustness of those relationships.

The squared change of inflation appears to be a robust determinant of mean inflation uncertainty, even after controlling for actual and expected inflation, lending some support to the fact that inflation uncertainty is to some extent related to inflation volatility. However, this seems to be less robust if one considers the variance of the aggregate distribution as measure of uncertainty, and even less so for disagreement, since after controlling for inflation or inflation expectations the relationship weakens or becomes insignificant. The volatility of inflation seems to have no effect in the asymmetry of the perceived inflation risks.

As regards economic activity variables, although inflation risks appear to have no relationship with the output gap measure, they instead are strongly influenced by the perceived uncertainty about real GDP growth and also to respond to the perceived probability of recession over the horizon of the inflation forecast. This not only reflects the role of the perceived state of the real economy when forming inflation expectations but also underscores the forward-looking nature of inflation expectations reported by SPF panelists.

Table 5: Inflation risks and macroeconomic variables

	Variance	Mean uncertainty	Disagreement	Skewness	Mean-Mode
Panel A: Bivariate regressions (each cell represents a separate regression)					
Inflation	0.19*	0.16**	0.05**	0.04	-0.04
Δinflation-squared	0.89*	0.63*	0.26*	0.41	0.13
Output gap	-0.82	-1.06	0.33	0.82	-3.20
Prob. recession	0.04**	0.04**	0.01**	0.04*	0.02*
Growth uncertainty	0.98**	0.83**	0.20**	1.06**	0.53**
Panel B: Controlling for actual inflation					
Δinflation-squared	-0.43	-0.41*	-0.04	0.34	0.31
Output gap	5.47	3.78	1.66	-0.57	-3.27
Prob. recession	0.04**	0.03**	0.01**	0.04*	0.02
Growth uncertainty	0.78**	0.70**	0.15**	1.26**	0.69**
Panel C: Controlling for expected inflation					
Δinflation-squared	-0.55*	-0.51**	-0.09	-0.34	-0.15
Output gap	4.94	3.47	1.57	1.62	-1.59
Prob. recession	0.02	0.02*	0.01	0.03	0.02
Growth uncertainty	0.65**	0.62**	0.10*	1.17**	0.63**

** and * denote significance at 1% and 5% level respectively (Newey-West corrected standard errors upto a year)

Prob. recession refers to the reported probability of negative real GDP growth 4 qtrs ahead in the SPF.

Growth uncertainty refers to the estimated mean uncertainty from the SPF probabilistic forecast.

In sum, the evidence collected in this section provides a number of interesting insights about the developments in perceived inflation risks over the last four decades:

- perceived inflation risks seemed to be much less volatile than central tendency measures of inflation expectations. In turn, inflation uncertainty appears to have been much less volatile than the asymmetry in the risks surrounding inflation expectations. Notwithstanding this a significant moderation is evident since the early 1990s
- in terms of comovement between moments, (mean) inflation expectations seem to be fairly strongly correlated with uncertainty measures but little with the measures of asymmetry in risks.
- there is supporting evidence that higher inflation has been usually associated with higher inflation uncertainty, as well as higher inflation volatility.
- the business cycle appears to have a strong influence on inflation risks, but that relationship seems to be forward-looking: inflation risks tend to decline ahead of recessions, and be strongly influenced by the perceived uncertainty about real GDP growth and also to respond to the perceived probability of recession over the horizon of the inflation forecast.
- outside recession periods since the early 1990s the impact of the business cycle on both inflation uncertainty and upside risks seems to have weakened significantly.

This being said, it however remains a need to assess the relevance of the additional information content the perceived uncertainty and risk assessment may convey. The next section provides an example.

5 Inflation risks and inflation scares in the 1980s

One of the aims of this paper is to illustrate that a combined analysis of movements in all three moments of inflation expectations analysed above may shed new light on two crucial issues for monetary policy: (i) on the dynamics of private sector´s inflation expectations in general; (ii) on some puzzling episodes in bond markets, via the potential relationship between perceived inflation risks and premia embodied in bond yields. This section explores that conjecture by focusing on the evolution of perceived inflation risks during the disinflation and stabilization of inflation that took place under former Fed chairman Volcker in the 1980s.

The Volcker disinflation is one of the most visible macroeconomic events of the last 50 years of U.S. history and has had an enormous influence on the theory and practice of monetary policy.[34] It was a decisive turning point in the postwar monetary history of the United States: inflation had been dramatically rising, but under chairman Volcker the Fed first contained and then reversed that process after a difficult period of sustained disinflationary monetary policy.

It is therefore not surprising that the Volcker disinflation period remains the subject of a significant amount of research and spirited discussion. Goodfriend and King [2005] has recently provided an excellent historical analysis of the Volcker disinflation and persuasively argued that one of the key features of the reduction in inflation engineered by the

[34] See Blinder [2004] and Fisher [1994] among others.

Volcker Fed was its imperfect credibility.[35] Indeed there seems to be widespread consensus that the principal cause for the economy's costly recovery was the lack of credibility of the disinflation.[36]

What is surprising is that a thorough analysis of the dynamics of inflation expectations in the Volcker disinflation period appears to remain missing. This paper aims at filling that void by analysing available survey data at that time. Despite the usefulness of long-term interest rates as indicators of inflation expectations, changes in inflation expectations are only one of the possible factors behind movements in nominal yields, so survey measures may offer a more direct source of information about the evolution of private sector´s inflation expectations over that period.

Our analysis of the dynamics of private sector´s beliefs about inflation allows us to interpret how credibility was acquired by the Volcker Fed. As other contributions in the analysis of the Volcker disinflation, our analytical approach focuses on the interplay between inflation, expected inflation, credibility and long-term interest rates. Modelling in detail the interaction between the Fed policy and private sector beliefs about inflation, although desirable, is however beyond the scope of this paper.[37] Nor do we provide a thorough historical account of all the events surrounding monetary policy decisions, as that can be found in earlier excellent contributions (see Goodfriend [1993] and Goodfriend and King [2005] among others). We instead focus on the original piece of our analysis, the dynamics of inflation expectations over that period, and cast our interpretation around the "inflation scares"—falling bond prices due to sharply rising inflation premia in long-term interest rates[38]—during the Volcker era. Under the assumption that our measures of perceived inflation risks and bond risk premia should be closely related, we believe it is more fruitful to combine information from both surveys and bond markets and interpret the inflation scares identified through long-term bond yields as particular periods along the process of acquiring the desired credibility for monetary policy.[39]

Our risk measures suggest that it took quite some time for the Volcker Fed to acquire that credibility. Moreover, the movements in the inflation risk measures derived in this paper around the time of the bond market inflation scares suggests a close link between the two, and a taxonomy of those inflation scares on the basis of direct evidence on private sector inflation beliefs emerges as a by-product of our interpretative historical analysis.

[35] Goodfriend and King [2005] historical analysis hinges on two important pieces of evidence extracted from the transcripts of the Federal Open Market Committee. First, compelling evidence that former chairman Volcker and other FOMC members thought that acquiring credibility for low inflation was central to the success of their disinflation. Second, that long-term interest rates were regarded as indicators of inflation expectations and of the credibility of the Fed´s disinflationary policy. Indeed, the behavior of long-term interest rates, which rose several times despite the absence of monetary policy tightening and the decrease in inflation is interpreted as indicating that financial markets expected high inflation to return.

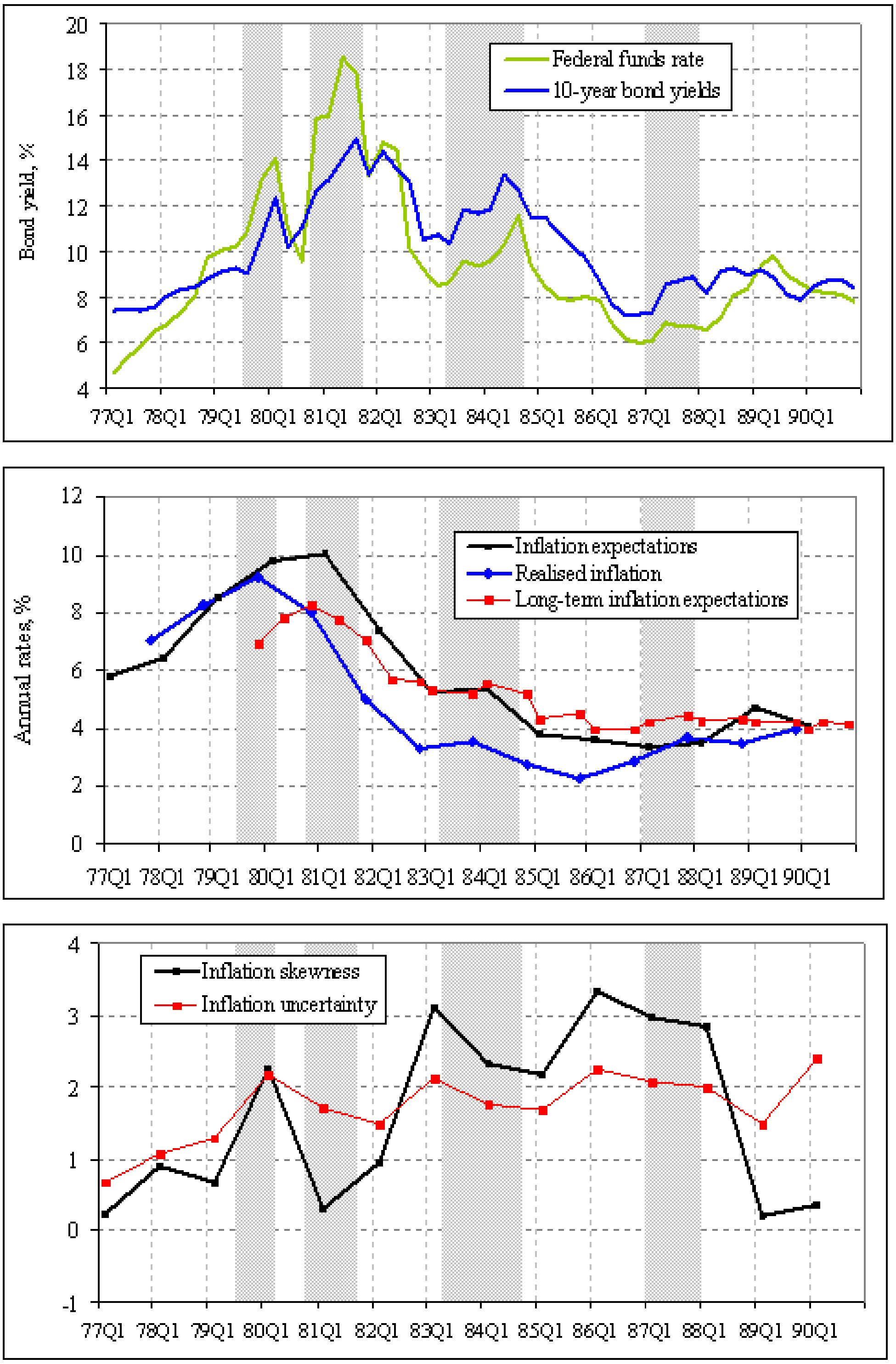
[36] The crucial role of credibility for a disinflationary process to cause output losses even in the presence of nominal rigidities was established by Ball [1994]. See also Ball [1995].

[37] For an analysis of learning and inflation scares see for example Orphanides and Williams [2005].

[38] See Goodfriend [1993], Gurkaynak, Sack, and Swanson [2005] and Orphanides and Williams [2005].

[39] It has to be borne in mind that the different frequency of the financial and survey data makes it difficult to match exactly their movements.

Figure 16: Interest rates, inflation expectations and inflation scares



Note: In the second chart inflation expectations denotes the mean of the aggregate distribution for the current year. Realised inflation denotes the year-end rate of growth in the GDP deflator and long-term inflation expectations is the combined series as reported by the Federal Reserve Bank of Philadelphia. The third chart depicts the variance and skewness estimated from the aggregate distribution.

Figure 16 (second chart) shows developments in actual and expected inflation, both over short-term (four-quarters ahead) and long-term horizons, as well as in inflation uncertainty and the asymmetry in risks embodied in the SPF probabilistic forecasts during the Volcker era. In order to focus our discussion, the periods usually associated to inflation scares in the bond market are represented by shadowed areas.

Several stylised facts emerge from those charts. First, looking at expected inflation as measured by the point estimates (mean expected inflation), while in the late 1970s they tended to be below realised inflation, they did catch-up towards the turn of the decade and exceeded its level during most of the 1980s, which lends support to the thesis of a very slow improvement in the Fed´s credibility. Indeed, available survey measures suggest that long-term inflation expectations declined in the first half of the 1980s but remained broadly unchanged for most of the second half, fluctuating above actual inflation rates and expected inflation over shorter horizons. Moreover, higher-order moments of inflation expectations also exhibited significant fluctuations in the 1980s, with upward movements in both inflation uncertainty and the perceived asymmetry of risks around the time of the inflation scares in the bond market. We analyse them in turn.

Early stages of the disinflation: the 1979-80 and the 1981 inflation scares

By the time Paul Volcker became Fed Chairman in August 1979, oil price developments greatly worsened the inflation outlook. The Fed rose rates on its October meeting, and the switch to non-borrowed reserve targeting officially reinforced the anti-inflation determination. The Fed however paused in its tightening at the end of year with the federal funds rate around 13.5% and the ten-year rate at 10.5%. The long-term rate then edged above 13% in February 1980, forcing the Fed to hike interest rates. These actions however only appeared to help contain inflation temporarily, but not to reduce it, while at the same time triggered a recession that halted the Fed tightening.

By late 1980, the ten-year bond rate had reversed the 2 percentage point decline associated with the recession and was back above 12%. Moreover, despite having the federal funds rate at 19% in 1981, the FOMC again faced a period of steadily rising long-term interest rates to a peak in excess of 15% in October 1981 after a 3 percentage point gain between January and October 1981. The second inflation scare occurred in spite of an even more determined tightening of interest rate policy than occurred in late 1979.

Figures 14 and 15 showed that inflation expectations worsened significantly over that period. (Mean) expected inflation steadily rose between 1979 and 1981, reaching two digits in 1981 despite the decline in realised inflation since 1980. Besides, inflation uncertainty rose sharply to a historical peak in 1980 to moderate somewhat later on. Indeed, the variance decomposition shown in Figure 9 suggests that rising inflation uncertainty was accompanied by a rise in both disagreement about the inflation outlook at the time and the individual uncertainty surrounding that outlook. Finally, perceived inflation risks were also tilted towards higher inflation outcomes as indicated by the upward shift in the skewness of the probabilistic forecasts.

The 1983-84 inflation scare

The Fed had taken a strong disinflationary policy since 1980, but two years later, with inflation coming gradually down, the Fed somewhat relaxed its policy stance: by February 1983, the monthly average funds rate was around 8.5% while the ten-year rates remained steady at around 10.5%. A few months later, however, with economic activity gaining momentum, the long-term rates started to rise fast, reaching almost 12% by end-1983 and peaking at around 13.5% in mid-1984, only about a percentage point short of its October 1981 peak, even though annualised (quarterly) inflation as measured by the GNP deflator was below 5% throughout most of 1983 and 1984, that is about 4 or 5 percentage points lower than in 1981. The Fed tightened in an effort to resist the on-going inflation scare, but long-term rates only starting decreasing in the second half of 1984 and did so very gradually, only returning below the 10% level in the last months of 1985. It has been argued that the successful containment of the 1983-84 inflation scare was arguably the most remarkable feature of the Volcker disinflation (Goodfriend [2005]). The evidence presented in this paper also suggest that the 1983-84 inflation scare was significantly different of the two previous ones in several crucial aspects.

Our analysis reveals somewhat surprising movements in inflation expectations around the time of the 1983-84 bond market inflation scare. Basically, inflation expectations (as measured by the mean of the consensus distributions) actually *declined* during most of the 1980s and particularly between the first quarters of 1983 and 1986 (by more that 1.5 percentage points, see Figure 16).[40] However, despite the decline in the (mean) inflation expectations, over the same period inflation uncertainty rose and inflation risks edged strongly upwards since early 1983 and remained high over most of the bond market scare period, most likely reflecting that private agents believed that the reduction in inflation was temporary, and a return to high inflation likely. In sharp contrast to the two previous inflation scares, the then imperfect credibility of the Fed appeared to be reflected not in the usually considered point estimates of inflation expectations but in the *risks* (uncertainty and in particular asymmetry) associated to the inflation forecasts, whose trend de-coupled from that of mean inflation expectations in the 1982-86 period.

The 1987 inflation scare

In 1987, the Volcker Fed was confronted with a fourth inflation scare. With inflation rising for the first time in the decade, the inflation scare was marked by a sharp but relatively brief 200-basis-point rise in long-term bond yields between March and October 1987. (Mean) inflation expectations reacted little to the rise in actual inflation rates, allowing for the gap between the two to close by early 1988, while longer term inflation expectations appear to move only slightly upwards around 1987. At the same time, perceived inflation risks, rose again from early 1986, and only declined significantly once the scare was contained.

Assessment

This evidence illustrates that changes in inflation expectations may be quite complex, and it is very likely that a full analysis of inflation risks characterised by higher-order moments of inflation expectations is necessary to correctly interpret them. In particular, the above analysis of the four inflation scares faced by the Volcker Fed in the 1980s, shows that such episodes may be triggered by significantly different movements in inflation expectations. While the first two inflation scares reflected a rise in all three moments of inflation expectations, movements in long-term rates around 1984 could have been triggered by changes in the perceived inflation risks, most likely affecting risk premia. Instead, the information provided exclusively by the point estimates is insufficient, if not

[40] Figure 13 depicts about four-quarters ahead inflation expectations consistent with the higher-order moments derived in this paper. However, available survey measures of long-term inflation expectations show a similar pattern (data are available through the webpage of Federal Reserve Bank of Philadelphia).

misleading, to explain this episode. Although the relationship between the higher-order moments of inflation expectations and the risk premia embodied in bond yields is likely to be rather complex, this section shows that a thorough analysis of the density function describing expected inflation may shed new light in this regard.

6 Concluding remarks

This paper has proposed a new methodology to extract information about the first three moments of the inflation density forecasts provided by the Survey of Professional Forecasters (SPF). Our approach comprises two methodological contributions to the related literature. First, a small departure from maximum likelihood as an efficient estimation approach. Second, the use of a parsimonious but nonetheless very flexible theoretical density function, the skewnormal. We have shown that both of those methodological contributions lead to substantial accuracy gains in the estimation of the moments of the SPF distributions of inflation expectations.

An application of this methodology to the actual SPF data leads us to conclude that asymmetries, and therefore asymmetries in risk assessment surrounding the baseline scenario, although often neglected so far, is an intrinsic feature of inflation expectations (at least at the horizons requested by the SPF). We have also shown that, moving beyond point estimates of inflation expectations, risks measures surrounding them are necessary to provide a thorough account of private sector's beliefs about the inflation outlook. Furthermore, we illustrate how our inflation risk measures shed some new light on key economic events, such us the inflation scares in the U.S. bond market in the mid 1980s in the context of the Volcker disinflation.

This evidence suggests that a thorough analysis of the probability distributions of the SPF may provide additional insights not only about how inflation expectations may change over time but also about developments in other macroeconomic and financial variables. We hope that the methodology developed in this paper and our analysis of the dynamics of inflation beliefs in the Volcker disinflation period could be a first step in that regard.

A. On the appropriate fitting criterion for the SPF histograms

A.1. Shortcomings of the (unweighted) least squares criterion

Recent literature that considers the pros and cons of fitting parametric densities to the SPF histograms has considered least squares as fitting criterion, i.e. minimising the sum of the squared deviations of the observed probabilities with respect to the theoretical ones over the set of intervals. More formally, $LS = \sum_{i=0}^{I} \left(\widehat{p}_i - p_i(\varrho)\right)^2$. We have argued that minimising the LS criterion (although consistent) is not efficient, for it does not use all available information in the problem at hand. Our probabilistic framework indeed allows to ground the choice of criterion on statistical inference arguments. Some additional insights can be gained by considering that generalised least squares (GLS), a standard (efficient) approach, would suggest minimising $GLS = (\widehat{\mathbf{p}} - \mathbf{p}(\varrho))' \, S^{-1} \, (\widehat{\mathbf{p}} - \mathbf{p}(\varrho))$ with respect to ϱ where S^{-1} denotes the inverse of the covariance matrix of the frequencies vector, $\widehat{\mathbf{p}}$. As outlined in the previous section, our multinomial framework allows for a direct identification of the structure of the matrix S without imposing any *ad-hoc* restriction. Specifically, the structure of the covariance in a multinomial model is well known to be of the form

$$S = \begin{pmatrix} p_0(1-p_0) & -p_0p_1 & & \cdots & -p_0p_I \\ -p_0p_1 & p_1(1-p_1) & & \cdots & -p_1p_I \\ \vdots & & \ddots & & \vdots \\ -p_0p_{I-1} & \cdots & & p_{I-1}(1-p_{I-1}) & -p_1p_{I-1} \\ -p_0p_I & \cdots & & \cdots & p_I(1-p_I) \end{pmatrix}$$

It can be seen from its structure that the matrix S is singular (as the sum of frequencies p_i must add up to one). Its Moore-Penrose inverse, S^-, has the following structure [41]

$$S^- = \begin{pmatrix} 1/p_0 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1/p_I \end{pmatrix}$$

It is then immediate that it suggests the use of the weighted least squares (WLS) estimator. In this case, minimising WLS is algebraically equivalent to minimise the Pearson chi-square criterion , defined as:

$$X^2 = \sum_{i=0}^{I} \frac{\left(\widehat{p}_i - p_i(\varrho)\right)^2}{p_i(\varrho)} \tag{5}$$

Alternatively, the unknown vector $\mathbf{p}(\varrho)$ could be replaced in the denominator by its estimate, $\widehat{\mathbf{p}}$ to obtain the modified chi square criterion due to Neyman [1949].

The expression above highlights a fundamental difference between LS and WLS: while the LS assigns equal weighting to the fitting errors for each interval, an efficient WLS assigns a weight proportional to the inverse of the observed (or fitted) frequency, as the best available estimator of the unobserved probability, by means of the denominators in

[41] Note that the structures of the S and the S^- matrices come directly from the interpretation of the reported histograms as the realisations of a multinomial variable and we are therefore not imposing any additional restriction on the their structure that could potentially lead to instability of the GLS (or WLS) estimators arising from the potential misspecification of those structures.

equation (5). If the frequencies were constant across the intervals i, LS and WLS would provide the same results. More formally, in terms of the above expressions, for WLS (or NM^2) and LS to be equivalent, the matrix S^- should be proportional to the identity matrix, thereby providing equal weighting for all the intervals. However, in the case of the SPF probability distributions, that is unlikely to be the case: the observed relative frequencies in each interval are far from similar, but rather resemble a strongly bell-shaped function. It thus looks very unlikely that assigning the same weight to the fitting errors at the center of the distribution and at the tails, as the LS criterion does, would make efficient use of all the information contained in the SPF data.

The above discussion highlights a fundamental shortcoming of the commonly-used LS criterion. Indeed, the fact that the chi-square criteria derived above provides consistent, asymptotically normal and efficient estimates for the vector of distribution parameters is crucial, for the purpose of the exercise is to recover the key moments of the true underlying subjective density forecasts. This result is guaranteed for all estimators of the family of minimum power divergence

A.2. The family of minimum power divergence estimators

Here we present in greater detail the main properties of the family of power divergence estimators, which encompasses, among others, the X^2 and NM^2 criteria discussed above. Here we focus on the properties that are more relevant for our purpose in this paper, but the general properties of this family of estimators are extensively analysed in Cressie and Read [1988].

The family of minimum power divergence estimators can be defined in general terms as the estimates obtained by minimising the expression

$$I^\tau(\widehat{p}, p) \;=\; \frac{1}{\tau(\tau+1)} \sum_{i=0}^{I} \widehat{p}_i \left[\left(\frac{\widehat{p}_i}{p_i(\varrho)} \right)^\tau - 1 \right] \tag{6}$$

with respect to ϱ. The family is indexed by the parameter $\tau \in \mathbb{R}$.

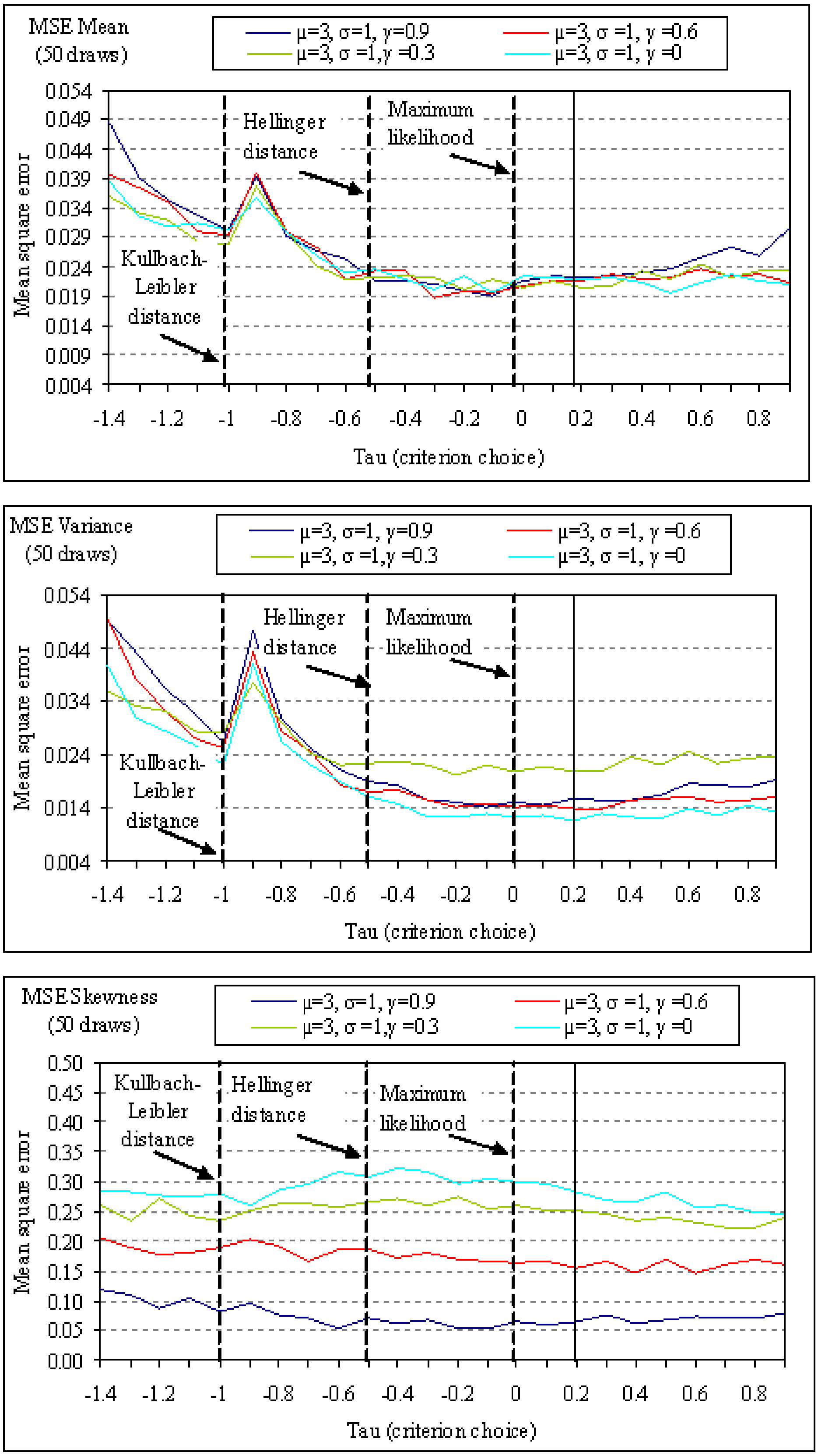
To grasp the broadness of this family, note that, if $\tau = -2$ for instance, equation (6) yields the Neyman Chi-Square criterion, while for $\tau = 1$ one obtains Pearson's X^2.[42]

Note that, again in this case, it is immediate from that expression that the efficiency gain with respect to the least squares criterion comes from the appropriate weighting of the fitting errors to account for the statistical properties of the data. Serfling [1980] shows that the estimate obtained from minimisation of the likelihood ratio G^2 (see equation (2)) is asymptotically equivalent in probability to the estimate obtained by minimising the standard Pearson chi-square criterion X^2 discussed above.

We have however argued that small sample properties are also fundamental in the choice of the optimal estimator in our problem. Indeed we searched for the optimal value of the parameter τ by examining estimation errors incurred for all three density parameters (μ, σ, γ). The Figure below shows the relative performance of some of the estimators discussed above based on the Monte Carlo experiment detailed in Section 3.2.

[42] Moreover, the family encompasses other widely-used estimators such as the Hellinger distance $\left(I^{-1/2} = 2\sum_{i=0}^{I}\left[\sqrt{\widehat{p}_i} - \sqrt{p_i(\varrho)}\right]^2\right)$ and the Kullback-Leibler divergence $\left(I^{-1} \equiv \lim_{\tau\to-1} I^\tau \;=\; \sum_{i=0}^{I}\left[p_i(\varrho)\log\frac{p_i(\varrho)}{\widehat{p}_i} + (\widehat{p}_i - p_i(\varrho))\right]\right)$.

Figure AI: Relative performance of alternative estimators, MSE



B. Moment decompositions

The third moment decomposition used in the main text is derived in as similar way to the standard ANOVA decomposition, as follows:

$$
\begin{aligned}
S(X) &= E\left[(X-E[X])^3\right] \\
&= E\left[E\left[(X-E[X])^3\,|Y\right]\right] && (7)\\
&= E\left[E\left[(X-E[X|Y]+E[X|Y]-E[X])^3\,|Y\right]\right] \\
&= E\left[E\left[(X-E[X|Y])^3\,|Y\right]\right] + \ 3\,E\left[(E[X|Y]-E[X])\,E\left[(X-E[X|Y])^2\,|Y\right]\right] \\
&\quad + 3\,E\left[(E[X|Y]-E[X])^2\,E\left[(X-E[X|Y])\,|Y\right]\right] \\
&\quad + E\left[E\left[(E[X|Y]-E[X])^3\,|Y\right]\right] && (8)\\
&= E\left[S\left(X|Y\right)\right] + \ S\left(E[X|Y]\right) + \ 3\,E\left[V\left(X|Y\right)\left(E[X|Y]-E[X]\right)\right] && (9)
\end{aligned}
$$

Step (7) is obtained applying iterated expectations. Step (8) follows from the expansion of the power and the measurability of $(E[X|Y]-E[X])$ with respect to the σ-algebra induced by Y. Finally, step (9) is obtained by observing that the third summand vanishes, since $(E[X|Y]-E[X])^2$ can be taken out of the expectation (due to its measurability) and the remaining factor is equal to zero.

C. Results from Monte Carlo analysis

The results reported in the columns in each of the tables below correspond to the power distance estimator (PDE) using a skew-normal density (PDE(τ^*), SN)), the power distance estimator using a normal density (PDE(τ^*), N), the standard least squares criterion using the Skew-normal density (LS (SN)) and the least squares criterion using the normal density (LS (N)).4 Hence, the first column represents our preferred methodology while the fourth represents the method most frequently used in the literature. The accuracy of the different estimators to recover the true underlying parameters of the underlying distribution is measured by computing mean square errors (MSE), mean absolute errors the errors (MAE), and the empirical bias obtained from 1000 simulated samples of the different sizes (i.e. a different number of random draws) in order to test the robustness of the proposed methodology.

Table AIII.1. DGP: *SN* (mean μ=3 , Standard deviation σ=1 , Skew γ=0.6)

		Mean μ				Standard deviation σ			
Sample size		PDE, SN	PDE, N	LS, SN	LS, N	PDE , SN	PDE, N	LS, SN	LS, N
n=20	MSE*	5.32	5.53	7.19	9.84	3.92	4.09	4.64	6.50
	MAE	0.19	0.19	0.22	0.26	0.16	0.16	0.18	0.21
	bias*	1.47	-0.26	-4.41	-13.2	-3.20	-1.88	-5.87	-12.3
n=50	MSE*	2.16	2.26	2.59	4.91	1.38	1.65	2.05	2.84
	MAE	0.12	0.12	0.13	0.18	0.09	0.10	0.12	0.14
	bias*	1.26	0.83	-1.99	-13.5	-1.35	0.44	-2.71	-8.12
n=100	MSE*	1.06	1.06	1.40	3.85	0.70	0.80	1.08	1.56
	MAE	0.08	0.08	0.10	0.16	0.07	0.07	0.08	0.10
	bias*	0.99	0.76	-0.82	-15.0	-0.53	1.82	-1.98	-6.15
n=200	MSE*	0.54	0.62	0.72	3.00	0.45	0.81	0.52	1.02
	MAE	0.06	0.06	0.07	0.15	0.06	0.07	0.06	0.08
	bias*	2.23	1.66	0.61	-15.0	2.55	5.07	-0.23	-6.31

Table AIII.2. DGP: *SN* (mean μ=3 , Standard deviation σ=1 , Skew γ=0.3)

		Mean μ				Standard deviation σ			
Sample size		PDE(, SN	PDE, N	LS, SN	LS, N	PDE, SN	PDE, N	LS, SN	LS, N
n=20	MSE*	5.74	5.46	6.21	8.11	3.00	3.61	4.57	5.62
	MAE	0.19	0.19	0.20	0.23	0.14	0.15	0.17	0.19
	bias*	1.29	-0.11	-1.46	-6.79	-3.08	-1.89	-3.45	-8.70
n=50	MSE*	2.07	2.33	2.90	3.79	1.40	1.39	1.86	2.26
	MAE	0.11	0.12	0.14	0.16	0.09	0.09	0.11	0.12
	bias*	0.49	-0.14	0.50	-6.81	-0.85	-0.13	-1.03	-4.98
n=100	MSE*	1.01	1.05	1.66	2.19	0.65	0.69	0.98	1.19
	MAE	0.08	0.08	0.10	0.12	0.07	0.07	0.08	0.09
	bias*	0.47	0.17	-0.36	-7.28	-0.70	0.26	0.16	-3.04
n=200	MSE*	0.61	0.56	0.78	1.41	0.47	0.57	0.48	0.60
	MAE	0.06	0.06	0.07	0.10	0.06	0.06	0.06	0.06
	bias*	0.95	1.26	0.61	-7.50	2.30	3.51	0.03	-2.43

Table AIII.3. DGP: *SN* (mean μ=3 , Standard deviation σ=1 , Skew γ=0)

		Mean μ				Standard deviation σ			
Sample size		PDE, SN	PDE, N	LS, SN	LS, N	PDE, SN	PDE, N	LS, SN	LS, N
n=20	MSE*	5.39	5.15	6.11	8.33	3.51	3.47	4.49	5.49
	MAE	0.18	0.18	0.20	0.23	0.15	0.15	0.17	0.20
	bias*	1.45	-0.53	0.70	0.18	-2.78	-1.58	-2.23	-5.89
n=50	MSE*	2.19	2.25	3.08	3.42	1.17	1.19	1.92	2.10
	MAE	0.12	0.12	0.14	0.15	0.09	0.09	0.11	0.11
	bias*	0.88	0.23	0.14	0.96	-1.26	-0.52	0.13	-3.11
n=100	MSE*	1.07	1.20	1.49	1.64	0.62	0.57	1.00	1.07
	MAE	0.08	0.09	0.10	0.10	0.06	0.06	0.08	0.08
	bias*	-0.02	0.38	-0.03	-0.32	-0.94	-0.24	1.10	-1.40
n=200	MSE*	0.60	0.56	0.82	0.90	0.41	0.46	0.52	0.52
	MAE	0.06	0.06	0.07	0.08	0.05	0.05	0.06	0.06
	bias*	0.50	-0.30	-0.06	-0.38	1.17	2.52	0.78	-0.98

*Reported figures are scaled up by 100.

References

Azzalini, A, "A class of distributions which includes the normal ones ", *Scandinavian Journal of Statistics*, 12, 171-178, 1985.

Ang, A., G. Bekaert, and M. Wei, "Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better?", *Journal of Monetary Economics forthcoming,* 2007.

Ball, L., "Credible Disinflation with Staggered Price-Setting." *American Economic Review*, March, 1994, 84, 282-89.

Ball, L., "Disinflation with imperfect credibility ", *Journal of Monetary Economics* 35, 5–23.

Beber, A., and M. W. Brandt, "The Effect of Macroeconomic News on Beliefs and Preferences: Evidence from the Options Market", *Journal of Monetary Economics,* Volume 53, Issue 8 , November 2006, Pages 1997-2039.

Birch, M. V. "A new proof of the Fisher-Pearson theorem", *Annals of Mathematical Statistics*, 35, 817-824, 1964.

Cressie, N. and T.R.C. Read. "Multinomial goodness-of-fit tests", *Journal of the Royal Statistical Society B*, 46, 440-464, 1984.

Cressie, N. and T.R.C. Read. "*Goodness-of-fit statistics for discrete multivariate data*". New York, Springer Verlag, 1988.

Croushore, D., "Introducing: The Survey of Professional Forecasters", Federal Reserve Bank of Philadelphia Business Review, November/December, 3-13, 1993.

D'Amico, S. and A. Orphanides, "Uncertainty and Disagreement in economic forecasting", Board of Governors of the Federal Reserve, *mimeo,* 2006.

Engelberg, J., C. F. Manski, and J. Williams, "Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters", NBER Working Paper #11978, Jan. 2006.

García, J.A. and A. Manzanares, "Comparing the information content of point estimates and probabilistic inflation forecasts", *mimeo,* 2007.

Goodfriend, M., "Interest Rate Policy and the Inflation Scare Problem:1979–1992", *Federal Reserve Bank of Richmond Economic Quarterly, Vol.* 79(1), Winter 1993.

Goodfriend, M. and R. G. King, "The incredible Volcker disinflation", *Journal of Monetary Economics* 52 (2005) 981–1015, 2005.

Giordani, P., and P. Söderlind, "Inflation Forecast Uncertainty", *European Economic Review*, 47, 1037-1059, 2003.

Henze, N. "A probabilistic representation of the 'skew-normal' distribution", *Scandinavian Journal of Statistics*, 13, 271-275, 1986.

Lahiri, K., C. Teigland, and M. Zaporowski. "Interest Rates and the Subjective Probability Distribution of Inflation Forecasts", Journal of Money, Credit and Banking, 20, 233-248, 1988.

Lindsay, B.G., "Efficiency versus robustness: the case for minimum Hellinger distance and related methods", *The Annals of Statistics*, 22, 1081-1114, 1994.

Machina M..and M. Rothschild, "Risk", The New Palgrave Dictionary of Economics, 2nd Edition, edited by S. N. Durlauf and L. E. Blume, Macmillan (Basingstoke and New York), *forthcoming.*

Neyman, J., "Contributions to the theory of the chi-square test", *First Berkeley Symposium on Mathematical Statistics and Probability*, 239-273, 1949.

Orphanides, "A. Comment on: “The incredible Volcker disinflation” ", *Journal of Monetary Economics* 52, 1017–1023, 2005.

Orphanides, A. and J. Williams, “Inflation Scares and Forecast-based Monetary Policy”, *Review of Economic Dynamics* 8, 498-527, 2005.

Patefield, M., and Tandy, D., "Fast and Accurate Computation of Owen's T-Function", *Journal of Statistical Software* Vol. 5, Issue 5, 2000.

Pearson, K.,"On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling", *Philosophy Magazine*, 50, 157-172, 1900.

Pewsey, A., "Problems of inference for Azzalini's skew-normal distribution", Journal of Applied Statistics, 27, 859-870, 2000.

Wallis, K. F., "Combining Density and Interval Forecasts: A Modest Proposal", *Oxford Bulletin of Economics and Statistics*, 67, 983-994, 2005.

Zarnowitz, V. and L. A. Llambros, "Consensus and Uncertainty in Economic Prediction", *Journal of Political Economy*, 95(3), 591-621, 1987.

Rao, C.R. "Efficient estimates and optimum inference procedures in large samples". *Journal of the Royal Statistical Society Series B*, 24, 46-72, 1962.

Rich, R. W. and J. S. Butler, "Disagreement as a Measure of Uncertainty: A Comment on Bomberger", *Journal of Money, Credit and Banking*, Vol. 30, No. 3, Part 1, 41 1-419, 1998.

Rich, R., and J. Tracy, "The Relationship between Expected Inflation, Disagreement, and Uncertainty: Evidence from Matched Point and Density Forecasts", Federal Reserve Bank of New York Staff Reports, No. 253, July 2006.

Serfling, R.J. "Approximation theorems of Mathematical Statistics", *J. Wiley. New York*, 163-165, 1980.

Witting, H. "Mathematische Statistik I", *B.G. Teubner. Stuttgart*, 330-331, 1985.