Are qualitative inflation expectations useful to predict inflation?*

Rolf Scheufele†

Halle Institute for Economic Research (IWH)

DRAFT
December 31, 2008

Abstract

This paper examines the properties of qualitative inflation expectations collected from economic experts. It describes their characteristics concerning rationality and Granger causality. Further, in an out-of-sample simulation study it is investigated whether this indicator is suitable for inflation forecasting. Results from other standard forecasting models are considered and compared with models employing survey measures. We find that a model using survey expectations outperforms most of the competing models. Moreover, we find some evidence that the survey indicator already contains information from other model types (e.g. Phillips curve models). However, the forecast quality may be further improved by fully taking into account information from some financial indicators.

JEL-Classification: C42, C52, E31

Keywords: Inflation expectations, Tests of rationality, Inflation forecasting, Forecast evaluation

*I would like to thank Henry Dannenberg, Wilfried Ehrenfeld and Dominik Maltritz for many helpful suggestions. I have also benefited from comments of seminar participants at the ifo workshop Macroeconomics and Business Cycles 2008 in Dresden and the IWH-CIREQ macroeconometric workshop 2008 in Halle. Special thanks to Tobias Kitlinski for excellent research assistance.

†Contact details: Halle Institute for Economic Research (IWH), Macroeconomics Department, P.O. Box 11 03 61, D-06017 Halle (Saale), e-mail: rse@iwh-halle.de, phone: +49 345 7753-728.

Contents

1 Introduction

Nowadays, inflation expectations play a central role for conducting monetary policy. Since many central banks have explicitly or implicitly adopted an inflation targeting regime, stabilizing inflation expectations have become the primary policy objective. Because there is a lag between policy actions and its impacts on the central banks target, monetary authorities are guided by forecasts. This makes inflation forecasting essential for effective monetary policy. Although monetary authorities seek to stabilize long-term inflation expectations, monitoring short- and medium-term inflation is important as well. Whenever inflation exhibits some inertia, good short term inflation forecasts translate into more-accurate longer-term projections.

In this paper we use survey data collected from economic experts and characterize its properties as indicators for German inflation expectations. The advantage of this indicator is its monthly availability and its fixed time horizon (6 months). We use a variant of the well known method of Carlson and Parkin (1975) to compute quantitative measures from qualitative responses. Further we test its properties concerning the assumption of rationality. Moreover, we investigate this measure of inflation expectations concerning its information content to predict future inflation rates (over a horizon of half a year). In an out-of-sample experiment we confront models that employ these survey measures with other inflation models. These other models consist of univariate time series models, Phillips curve specifications and term structure models. With tests for predictive accuracy we are then able to assess whether these models display significant different relative predictive accuracy.

The first contribution of this study is that it compares a broad range of popular forecasting models for inflation for Germany on a monthly base. Since Ang, Bekaert and Wei (2007) document the superiority of survey based methods over many alternative inflation models, we follow this line of research and investigate whether these results also hold for qualitative survey data for the German economy. Our second contribution is to get some insights about the expectation formation of economic experts. Here, the idea is to find out which information and which forecasting models are used by participants of the survey and which they do not.

Our findings can be briefly summarized as follows. Inflation expectations obtained from the ZEW Financial Market Survey are inconsistent with rational expectations, since they do not contain all available (costless) information and thus violate the orthogonality assumption. But a Granger causality test reveals that this series contains information about future inflation. In an out-of-sample experiment running from 2000.9 to 2008.7 we find that the pure survey measure performs poorly compared to other inflation models. Once an augmented model is used that includes not only the expected inflation series but also additional lags of actual inflation, the model outperforms most of the alternative specifications in terms of root mean squared error (RMSE). It does significantly better than a random walk model, an augmented term structure model and a benchmark autoregressive model. Only one model displays a lower RMSE (although the differences are not significant). Encompassing tests indicate that the survey model adjusted by serial correlation already contains the information of univariate time series models and Phillips curve specifications but disregards some information included in financial variables such as the interest rate or term spreads.

The remainder of this paper is as follows. Section 2 describes the data set and the conversion method to get quantitative inflation expectations. Section 3 presents the characteristics and some test results about the data. In Section 4 the out-of-sample set up is explained and its main results are presented. Section 5 concludes.

2 Measurement of Expectations via Survey Data

The use of survey data on inflation expectations has a long tradition in economic literature (e.g. Anderson, 1952; Theil, 1952). Those direct measures of expectations allows one to analyse the expectation formation process without relying on a particular behavioral model (which is typically done in rational expectation models).

In principle, it is possible to distinguish between two types of survey data on inflation expectations: "quantitative" and "qualitative". Quantitative means that respondents are asked for the exact magnitude of change or level. A question may be for instance "what inflation rate do you expect next year". In contrast to these exact measures, surveys may also ask for a general tendency. Here respondents give a qualitative statement, for example "do you expect that inflation goes up (or down) during next year". Although it seems always preferable to obtain point forecasts of expectations on future inflation rates, there may be also some drawback in using quantitative responses, because these direct measures may be rather affected by sampling and measurement errors compared to tendency statements (e.g. Pesaran, 1987, Ch. 8.2). Using qualitative survey data as measure of inflation expectations always requires to transform these data into quantitative expressions that requires many assumptions (which are generally not testable).

Another distinction between different survey types can be made according to the population of the survey. It may contain households, firms or professional economists. By far the two most popular survey data on inflation expectations in Europe are the EC Consumer Survey and the Consensus Forecasts (see Mestre, 2007, ANNEX A, for a detailed description for both surveys.).[1] The first

[1]Both surveys contain information not only for the Euro area as single area, but also on

survey contains monthly information about the personal economic situation of households including their inflation perceptions. The survey questions are of qualitative nature and ask about the past consumer price development as well as about the expected development of consumer prices during the next 12 month. The Consensus Forecasts consists of a panel of professional forecasters mainly from banks and economic research institutes.[2] Although the Consensus Forecasts are collected on a monthly basis, their usefulness is limited by the fact that forecasts do not have a fixed forecasting horizon since participants are always asked for the inflation rate for the actual year as well as for the upcoming year. So one alternative is to treat the surveys separately according to the month they were carried out. The other alternative is to stick on additional information from the quarterly consensus forecast where the expected development of consumer prices are reported for each single quarter. Each alternative is associated with a sizable reduction of the sample size.

2.1 Data set

In the paper we use a monthly survey that is carried out by the Center of European Economic Research (ZEW) to construct a direct measure for inflation expectations. This type of survey is more country specific than the ones mentioned above because only German financial analysts are consulted. This may be the primary reason why this type of survey plays only a marginal role in the literature about inflation expectations.[3] The ZEW Financial Market Survey covers about 300 experts from banks, insurances and large industrial firms. Each month, the experts are asked whether they expect "a rise", "a decline" or "no change" of the annual inflation rate in the medium term (during the next 6 months).[4]

The basic advantage of this data set is that it can be used to construct a monthly measure of inflation expectations with a fixed forecasting interval for each point in time. This is the main advantage over inflation expectations from the consensus economics forecast. A further advantage of the ZEW survey is its better representativeness since the number of participants is approximately 10 times higher compared to consensus forecast. Instead, a possible disadvantage of the ZEW survey is its qualitative nature. Here it is necessary to make additional

the some of the member states including Germany. Alternative surveys constitute industry surveys from the ifo-institute or from the European Commission and the Survey of Professional Forecasters of the ECB.

[2]For Germany the consensus panel consists of 30 institutions.

[3]Notable exceptions are Franz (2005) and Heinemann and Ullrich (2006).

[4]There is also an additional category of "don't know". Since this category in this particular survey is quantitatively unimportant this category is combined with the "no change" group. The questionnaire can be found under ZEW homepage (www.zew.de).

assumption to construct a quantitative expected inflation rate. This disadvantage is accepted here since there is a long tradition that deal with this caveat.

Compared to the available consumer surveys the ZEW survey may be less representative given the EU Consumer Survey for Germany which has a sample size of 2500, but also with respect to the attributes of the participants. However, this need not always be an advantage since it has been shown that consumers are quite heterogeneous in the perception of inflation. It has been found that people with lower educational attainment have problems to assess future price developments (see e.g. Kilian, Inoue and Kiraz, 2006). Further, we expect that professionals process information faster than consumers. So we expect that this survey type provides a better indicator for forecasting purpose.

2.2 Estimating German inflation expectations

The quantification of expected inflation rates is frequently based upon the probability method first proposed by Theil (1952). We follow this practice and employ a variant of the Carlson-Parkin method (Carlson and Parkin, 1975). This procedure rests on the idea that aggregate fractions of survey answers tell something about the magnitude of inflation changes. This makes it possible to get an aggregate measure of inflation expectations. The Carlson-Parkin-Method is the dominant procedure for quantification of qualitative inflation expectations in the literature.[5]

The quantification method requires some assumptions to hold. According to Pesaran (1987), these include:

- There exists an interval $[a_{it}, b_{it}]$ of inflation changes about zero which respondents cannot distinguish from zero.

- The subjective probability distributions have such properties that it is possible to obtain an aggregate probability distribution with first- and second order moments where the subjective information set is the union of the individual sets and where the aggregate expected change of the inflation rate is the average of the subjective expected change of the inflation rate.

- The thresholds a_{it} and b_{it} are the same across individuals, constant over time and symmetric around zero.

- The subjective probability distributions are independent of each other and of the same known form across respondents.

[5]Other possibilities are not feasible. For instance the regression approach of Pesaran (1985) or an augmented procedure of the Carlson-Parkin-Method introduced by Berk (1999) is not applicable because this data set contains no information on inflation perceptions of the past.

While the original Carlson and Parkin approach is employed for price levels, not for inflation rates, we have to modify their approach slightly that it fits to this particular data set. Following Carlson and Parkin (1975) we assume a standard normal distribution for the probability function of the inflation changes.[6] Further we estimate the scaling parameter $\widehat{\delta}$ in such a way that the series of changes of the expected inflation rate is on average unbiased. For the derivation of the expected changes of inflation we need the fraction of responses that report "inflation goes up" and the fraction of responses who say "inflation goes down" which we denote with A_t and B_t, respectively. We define $a_t = \Phi^{-1}(1 - A_t)$ and $b_t = \Phi^{-1}(B_t)$, where $\Phi^{-1}(\cdot)$ is the inverse of the probability function of the standard normal distribution. Now, we can calculate the expected change of the inflation rate during the next 6 months, $E\left(\Delta\pi^{12}_{t,t+6}|\Omega_t\right)$, with Ω_t the information set at time t and π^{12}_t defined as $\pi^{12}_t = 100\ln\left(P_t/P_{t-12}\right)$, the annual seasonally unadjusted inflation rate. This is a function of the variables a_t, b_t and δ_t given by

$$E\left(\Delta\pi^{12}_{t,t+6}|\Omega_t\right) = E_t\left(\Delta\pi^{12}_{t,t+6}\right) = -\delta_t\left(\frac{a_t + b_t}{a_t - b_t}\right). \tag{1}$$

To calculate this expression we need the survey data, the distribution function of the expected inflation changes and the value of δ_t. While the first two issues were already explained, it is important to emphasis the calculation of δ_t. As said above, we follow Carlson and Parkin (1975) and assume that, on average, expected changes in the inflation rates are unbiased and constant over time – with $\delta = \delta_t$ for every t.[7] With this assumption the scaling parameter can be calculated as

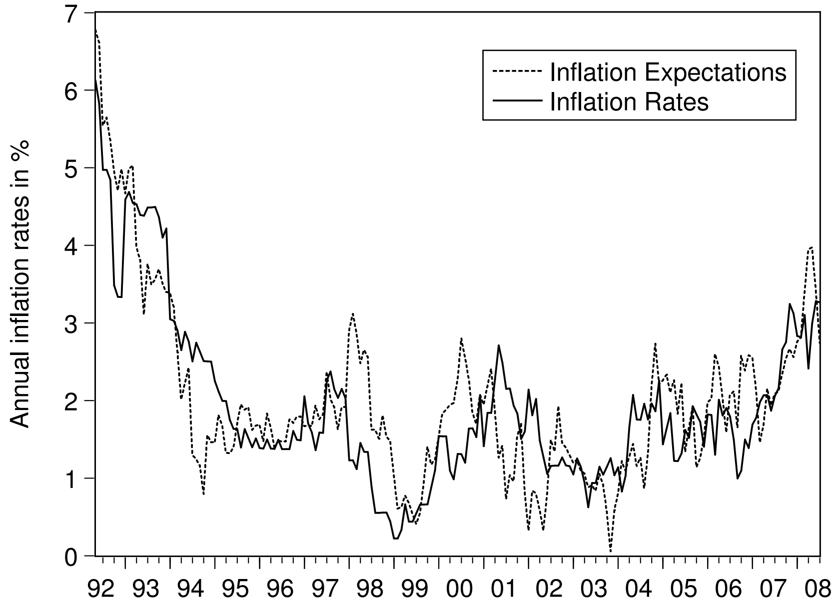
$$\widehat{\delta} = \frac{1}{T}\sum_{t=1}^{T}\left|\Delta\pi^{12}_{t,t+6}\right| / \frac{1}{T}\sum_{t=1}^{T}\left(\frac{a_t + b_t}{a_t - b_t}\right). \tag{2}$$

After the calculation of expected inflation changes it is also possible to compute expected inflation levels for the annual inflation rate. Since the actual inflation rate of the ongoing month is unknown at that point when the ZEW survey is carried out, expects base the expected change of the inflation rate on the most recent inflation release that refers to the inflation rate on the previous month.The

[6] According to Berk (1999) it makes little differences which particular probability distributions is used (other choices are the uniform distribution or the logistic distribution)

[7] We test whether inflation changes are asynchrony – implying that increases in the inflation rate are on average higher or lower than decreases. But a standard test of equal means indicates that increases and decreases in the inflation rate are on average the same.

Figure 1: Expected and realized inflation rates



expected inflation rate is then given by

$$E\left(\pi_{t+6}^{12}|\Omega_t\right) = E_t\pi_{t+6}^{12} = \pi_{t-1}^{12} + E_t\left(\Delta\pi_{t,t+6}^{12}\right) \quad (3)$$

where $E_t\pi_{t+6}^{12}$ is the expected annual inflation rate for time $t+6$ and π_{t-1} is the inflation rate to which experts base their opinion about changes in the inflation rate (that is known to the survey participants).[8]

3 Some Basic Properties and Tests

Since we want to explore whether the survey of financial expects provides useful information about future inflation rates we first describe the properties of the expected inflation series obtained in section 2. Our sample includes monthly annual inflation rates and expected inflation rates (over the next 6 months) from 1992.5 to 2008.7. Figure 2.2 shows how both series evolve over time. At the beginning of our sample both series came down from a very high starting point (after German reunification). Then, they fluctuate around a constant trend, slightly below two percent. Since it is still an open question whether inflation rates are characterized by stationary behavior or whether they are integrated of order one, we employ unit

[8]We use the seasonally unadjusted CPI inflation rate to construct the inflation series. This inflation indicator receives most attention in the German public. We also used its seasonal adjusted counterpart, but we found that the unadjusted series perform better.

Table 1: Test for (Granger-) causality between inflation and expected inflation based on VAR(9) model

Causality hypothesis	Test value	Distribution	p-value
Inflation $\overset{Gr}{\rightarrow}$ Inflation expectations	62.88	$F(9, 183)$	0.00
Inflation expectations $\overset{Gr}{\rightarrow}$ Inflation	4.55	$F(9, 183)$	0.00

Note: Sample period: 1993M03-2008M05. Lag length is selected due to the AIC.

root tests to explore whether annual inflation rate based on the monthly consumer price deflator and the expectation series contain a unit root. We use two standard tests, namely the Augmented Dickey-Fuller (ADF) test and the Phillips-Perron (PP) test (see Appendix A.1 for results), to test for a unit root. Both tests indicate that inflation and inflation expectations are stationary in levels.

The literature on measured inflation expectations apply various tests to investigate whether expectations are formed rationally. These tests can be characterized by 4 hypothesis associated with the rational expectation assumption. The test hypothesis may be based on unbiasedness, lack of serial correlation, efficiency and orthogonality (Pesaran, 1987). Since our approach for quantification of inflation expectations already includes more or less the assumption of unbiasedness, so it makes little sense to test for this specific assumption. Moreover, we also have to account for measurement errors in our series of inflation expectations. This is an additional argument why tests of unbiasedness and lack of serial correlation are not appropriate in our setting. For instance, the frequently applied test of unbiasedness by regressing the inflation series on expected inflation is very difficult to interpret when we assume measurement errors in the expected series. In this case the test is based on the null hypothesis that the constant is zero and the coefficient of the expected series is equal to one. It is clear that when we allow for measurement errors the coefficient of the expected series is biased toward zero. So whenever the test of unbiasedness rejects it is not clear whether this is due to measurement errors or due to a lack of unbiasedness. Nevertheless, we also report these standard tests but their results should be treated with caution. The traditional static test of unbiasedness rejects for the 5% significance level (see A.2), but we find it hard to interpret this result due to the problems adressed earlier.

Additionally, we also present some evidence about the direction of causality within a standard Granger test (see Table 1). This causality test helps determine whether the expected inflation series contains additional information about future inflation beyond what is already contained in the past history of actual inflation. If inflation expectations do not predict inflation, it will be a bad indicator of inflation. In this case, we find a feedback effect between inflation and inflation expectations –

this stands in contrast to some other studies that identify a unidirectional causality running from inflation to inflation expectations (Berk, 1999; Forsells and Kenny, 2004). This result indicates that our series of inflation expectation is suited for predicting future inflation[9] - a result that is also important for monetary policy.

Our emphasis concerning the rationality assumption is on testing the orthogonality assumption. The basic idea is to check whether the forecast error $\pi_t^{12} - E_{t-6}\pi_t^{12}$ is orthogonal to the (costless) information set available at the time when expectations are formed. If this hypotheses cannot be rejected this would indicate that expectations are formed in a rational way. It should be clear that it is only possible to use a very small subset of this general information set. We follow the existing literature (e.g. Pesaran, 1987) and consider macroeconomic variables to proxy, at least partly, the available information set. Basically, we can categorize our variables under consideration into three classes of variables: Variables that reflect real economic activity (industrial production, unemployment rate), financial variables (short and long term interest rates as well as the spread between the two) and a price variable reflecting price changes of foreign raw materials and energy. We consider all variables with lags up to one year after the time when expectations are formed.[10] For industrial production and unemployment rate we additionally construct detrended measures using a quasi real-time HP-filter (that uses only past information). For other variables we take year-on-year differences (see also section A.4 for an exact variable definition).

The orthogonality test is conducted in the following way. We run regression

$$\pi_{t+6}^{12} - E_t\pi_{t+6}^{12} = c + x_t'\beta + e_t, \tag{4}$$

where x_t contains all relevant explanatory variables including lags of inflation and the variables mentioned above and c as a constant. We exclude all insignificant variables in a stepwise procedure and obtain the final model by minimizing the SIC. Then we test the joint null hypothesis that $c = 0$ and $\beta = 0$. Since autocorrelation and heteroskedasticity may be an issue in this model we use Newey-West standard errors to calculate the test statistic. Table 2 shows the results of two different specifications. The first considers all explanatory variables together with lagged inflation terms. In the second specification we omitted the lagged inflation terms.

In the first specification eight variables turn out to be significant and explain the difference between inflation and inflation expectations. These are mainly financial variables like the short term interest rate and the spread between long-term and short-term interest rates. Further, the unemployment gap and the price changes of commodity prices seem to matter for the first specification. The joint test is highly

[9] This is confirmed by a static test of forecastability (see section A.2)

[10] Due to publication lags we use only variables that are available at time t when expectations are formed.

Table 2: Orthogonality Test

Dependent Variable: $\pi^{12}_{t+6} - E_t\pi^{12}_{t+6}$	I		II	
Variable	*coef*	*se*	*coef*	*se*
Constant	1.447	(0.22)	0.022	(0.12)
π^{12}_{t-1}	−0.725	(0.12)	–	–
π^{12}_{t-2}	0.219	(0.11)	–	–
$\Delta_{12}r^{s}_{-1}$	–	–	−0.202	(0.06)
$\Delta_{12}r^{s}_{-12}$	−0.333	(0.07)	−0.106	(0.06)
$\left(r^l - r^s\right)_{t-7}$	−0.438	(0.11)	–	–
$\left(r^l - r^s\right)_{t-9}$	–	–	−0.230	(0.07)
$\left(r^l - r^s\right)_{t-12}$	−0.209	(0.11)	–	–
$\Delta_{12}p^{raw}_{t-12}$	2.271	(0.28)	–	–
$Ugap_{t-1}$	1.255	(0.12)	–	–
$Ugap_{t-3}$	−0.602	(0.29)	–	–
$\overline{R}^2$	0.56		0.31	
SIC	1.54		1.90	
χ^2-Test	146.2***		49.8***	

Variable selection is done in a step-wise procedure minimizing SIC. The χ^2-Test is the test of joint significance of all coefficients. Newey and West standard errors in brackets. ***: 1% significance level. Sample period: 1992.5-2008.2.

significant, irrespectively whether one include lagged inflation terms or not. This implies that the rational expectation hypothesis can be rejected for this particular data set, which corresponds to many other studies that test the orthogonality assumption for surveys (e.g. Forsells and Kenny, 2004). This result implies that respondents can improve their inflation valuation by using the available information more efficiently.

4 Survey Expectations as Indicator for Future Inflation

In this section we conduct an out-of-sample forecasting experiment with different competing models for inflation forecasting. The main attention rests on the relative forecasting performance of models that employ survey-based measures as introduced in the previous sections. Three alternative model classes for inflation forecasting in Germany are examined: univariate time series models (autoregressive and unit root models), regressions with real activity variables (motivated from the Phillips curve) and term structure models.

4.1 Forecasting models and setup

The ability of inflation forecasting plays a crucial role for monetary policy in conducting optimal monetary policy. Further it is important for private agents for price setting, optimal investment or negotiation of wage contracts. Here we ask whether qualitative inflation expectations may contribute to standard inflation models in terms of forecasting performance. By standard inflation models we mean easy applicable statistical models that were found to be good models for predicting inflation in the literature. So the idea is that inflation expectations from surveys (in particular those from economic experts) should contain lots of information that is also present in single equation models with a few number of explanatory variables. Ang et al. (2007) document the forecasting performance of different surveys for inflation expectations in the US and their relative forecast accuracy. They find that survey measures provide indeed important information to forecast future inflation. For Germany, no comprehensive study is available that assesses survey measures for predicting inflation.

We follow Stock and Watson (1999) and Ang et al. (2007) and conduct an out-of-sample experiment to assess the relative predictive power of the ZEW survey measure for inflation. This is the dominate model evaluation method because it is not subjected to overfitting in contrast to in-sample measures of model accuracy. As a natural benchmark we consider univariate time-series models of the ARIMA model class. We model German inflation rates as a stationary process. Since our test results (see table 4) are in line with this assumption. Moreover, it is also in line with New Keynesian models when the monetary authority follows a stable inflation objective (which seems plausible since the Bundesbank has provided a credible monetary policy regime over many years and was then replaced by the ECB with a similar objective).

Most of the forecasting models employed here consist of the following structure

$$\pi_{t+6}^{6} = \alpha + \beta\left(L^{6}\right)\pi_{t}^{6} + \gamma\left(L\right)x_{t} + u_{t+6}, \tag{5}$$

whereas π_t^6 is defined as $\pi_t^6 = 100\ln\left(P_t/P_{t-6}\right)$ with P_t the seasonally unadjusted CPI-Index in period t. x_t may include other explanatory variables such as real activity measures, financial indicators, expected inflation rates as well as seasonal dummy variables. To keep this setting more realistic, we only consider indicators for x_t that are available when forecasts are conducted. $\beta\left(L^6\right)$ and $\gamma(L)$ denote lag polynomials which are defined as $\beta\left(L^6\right) = \left(1 - L^6 - L^{12} - \ldots\right)$ and $\gamma(L) = \left(1 - L - L^2 - \ldots\right)$. This step-up implies a direct approach for multistep forecasting which is comparable with Stock and Watson (1999) and the natural setting for autoregressive distributed lag models that use the available information to form forecasts.

Next we describe the forecasting models that are used in the following out-of-sample forecasting experiment. First, we consider a simple univariate autoregressive model which serves as a first benchmark for all other models under consideration. This specification is a special case of Eq(5) where other regressors, except lagged inflation, are excluded. The lag length of this AR model is selected using the Schwartz criterion (SIC) for the in-sample period.[11] It turns out that the AR(2) model produce the best in sample fit according to the SIC. We estimate this model (as well as all other models) with and without additional seasonal dummy variables. Additionally another univariate time series model is considered – a random walk model (RW) – inspired by Atkeson and Ohanian (2001). These authors find out that many inflation models for the US (including Phillips curve models) are not able to outperform a simple unit root model. This result is confirmed by Fisher, Liu and Zhou (2002) who document the superiority of random walk forecasts for inflation in the nineties over other models of inflation. The unit root model is constructed as a yearly average of past inflation rates defined as

$$\pi_{t+6}^{6} = 1/2\left(\pi_t^6 + \pi_{t-6}^6\right).$$

The second model class consists of Phillips curve models. By Phillips curve models we mean specifications that link the inflation rate with some measure of real activity like the unemployment rate or an output gap variable. Still, the Phillips curve plays a prominent role in theoretical monetary models that are known as New Keynesian models (see Galí and Gertler, 1999, for the empirical success of the New Phillips curve). We do not estimate these structural models here, instead we stick to a reduced form model that can be seen as a rough approximation of a more sophisticated Phillips curve model.[12] Different real activity measures are considered (and included for x_t in the general specification of Eq(5)): the unemployment rate (which is the real activity measure in early Phillips curve specifications when a constant NAIRU is assumed), the deviation of the unemployment rate from a quasi real-time Hodrick-Prescott-Trend (motivated from a time-varying NAIRU assumption) and detrended industrial production (also quasi real-time detrended using the HP-filter which corresponds to an output gap measure) together with changes in industrial production. Again the exact specifications with the particular lag are chosen according the SIC together with additional lagged inflation rates

[11] Stock and Watson (1999) consider a slightly different AR model specification. Their model consists of $\pi_{t+6}^6 = \alpha + \delta\left(L^1\right)\pi_t^1$, whereas our specification is equal to $\pi_{t+6}^6 = \alpha + \beta\left(L^6\right)\pi_t^6$. We choose this specification since monthly autoregressive terms (used by Stock and Watson, 1999) turn out to be instable and produce a very poor in-sample fit. Whereas using half-year inflation rates match the inflation dynamics better. We also stick to this autoregressive specifications for the following models.

[12] This corresponds to Stock and Watson (1999) and Ang et al. (2007) who use similar Phillips curve specification for their forecasting experiment.

(see table 6 for details).

Next, financial variables are used to predict inflation. Motivated by Fama (1975) and Mishkin (1990) we consider the short term interest rate as well as a term spread as inflation indicators. Several theoretical hypotheses rest on the idea that the yield curve is forward looking and may thus provide information about future inflation (e.g. see Kozicki, 1997, for theoretical arguments). We employ different models that include the short interest rate or the term spread or both as regressors for x_t in Eq(5). First, both measures – short term rate as well as the spread - are used as single explanatory variable (with additional lags of inflation). Second, we use both measures in one model and then we further augment Phillips curve specifications (as discussed above) with these variables (see table 6).

Finally, we consider survey indicators of expected inflation (in the definition as outlined in section 2) as inflation predictors. We begin by using directly the variable as constructed in section 2 although in a modified way. We compute the expectation series in such a way that we can use it to predict π^6_{t+6} with $E_t\pi^6_{t+6} = E_t\pi^{12}_{t+6} - \pi^6_t$. In a second specification we use $E_t\pi^6_{t+6}$ as a regressor together with a constant. Additionally, we use this measure as explanatory variable in Eq(5) by combining it with lagged inflation rates. This specification is motivated by the findings in section (3) where orthogonality tests show that expected inflation rates do not account for all autocorrelation present in observed inflation rates.

All models are estimated with the specific selection of regressors for the sample period 1992.5 to 2000.3. Then we obtain out-of-sample forecast with a fixed rolling window (2000.9 to 2008.7). For the remaining time period parameters are sequentially updated and we get forecasts for inflation π^6_{t+6} with a constant in-sample period. So we have always a fixed in-sample estimation period using 93 observations ($R = 93$) and we obtain exactly the same number of out-of-sample estimates for inflation ($P = 93$). We take the rolling scheme because this may guard against moment or parameter drifts which are very difficult to model explicitly (see, for instance, West, 2006, Section 4, for a discussion).

4.2 Forecast Evaluation

For examining the predictive accuracy of different models we assume a symmetric loss function given by mean square error (MSE) loss. Given this loss function, we may now evaluate and compare the outcomes of different models. In particular, we would like to assess how well models that employ survey measure of inflation expectations perform compared to other standard models for inflation. First, models are evaluated concerning their root mean squared error. Then we ask whether these differences are significant (which can be done by tests of equal predictive accuracy). Next we investigate the issue of forecast encompassing. The ideas is to examine whether forecasts from other models contain additional information not included in the reference forecast (irrespectively of its own forecasting performance in terms of RMSE).

Throughout in the evaluation step we consider parameter uncertainty due to estimation which is relevant for the specific tests. In the case of non-nested models we can conduct the Diebold-Mariano-Test (1996) to compare pairwise mean squared prediction errors. Let

$$\widehat{dif}_t^{(m,n)} = (\hat{e}_t^m)^2 - (\hat{e}_t^n)^2$$

be the standard square loss of model m relative to the loss of model n. A significance test in this case can be employed on the statistic

$$DM_6^{(m,n)} = \frac{1/93 \sum_{t=1}^{93} \widehat{dif}_t^{(m,n)}}{\widehat{\sigma}(\widehat{dif}_t^{(m,n)})} \overset{a}{\sim} N(0,1), \tag{6}$$

where $\widehat{\sigma}(\widehat{dif}_t^{(m,n)})$ is computed as a heteroskedasticity and autocorrelation consistent (HAC) estimate of the standard deviation.

For nested models we use an adjusted test as proposed by Clark and West (2006; 2007).[13] Additionally, we employ the testing methodology introduced by Giacomini and White (2006) that can be applied to nested as well as to non-nested models. This approach relies on conditional rather than an unconditional test and is recommended particularly for the rolling scheme as employed here. The test statistic is of a Wald-type and can be formulated as

$$GW_6^{(m,n)} = (93-6)\left(\frac{1}{(93-6)}\sum_{t=1}^{(93-6)} h_t \Delta L_{t+6}\right)' \hat{\Omega}^{-1} \left(\frac{1}{(93-6)}\sum_{t=1}^{(93-6)} h_t \Delta L_{t+6}\right), \tag{7}$$

with h_t a $q \times 1$ measurable test function which we set equal to $h_t = [1\ \Delta L_t]$ and a HAC covariance matrix $\hat{\Omega}$. Under some regularity conditions $GW_6^{(m,n)} \overset{a}{\sim} \chi_q^2$. ΔL_t refers to a loss function which is equal to $\widehat{dif}_t^{(m,n)}$ under mean square error loss.

To test for forecasting encompassing we rely on Harvey, Leybourne and Newbold's (1998) encompassing test, that can be formulated as

$$\hat{e}_t^m = \alpha + \delta\,(\hat{e}_t^m - \hat{e}_t^n) + \textit{seasonal dummies}, \tag{8}$$

with $\hat{e}_t^m$ the forecasting error of the benchmark model m and $\hat{e}_t^n$ the error of the competing model n. A standard t-test (along NW standard errors) can be used to

[13] They call the result MSPE-adjusted which can be computed as $1/93 \sum_{t=1}^{93} (\hat{e}_t^m)^2 - \left[\sum_{t=1}^{93} (\hat{e}_t^n)^2 - \sum_{t=1}^{93} (\hat{y}^m - \hat{y}^n)^2\right]$.

test for $\delta = 0$. Due to parameter uncertainty we follow West (2006) and adjust the test by a constant factor.[14] Further, the test is extended by additional seasonal dummies because some models contain seasonal dummies others not. So we intent to rule out test rejections that are simply caused by seasonal patterns.

4.3 Results

Table 3 reports the main results of our forecasting experiment. In terms of RMSE, defined in annual terms, only one model that employs survey inflation expectations is able to beat the simple AR model. The ZEW3 model that uses lagged inflation terms as additional regressors displays not only a smaller RMSE compared to simple time series models (AR and RW), but also relative to all Phillips curve models. Among the term structure models there is only one model (TS2) that displays a lower RMSE. So the pure survey measure displays a bad forecasting performance, but when we further adjust this model (by considering the bias as well as omitted autocorrelation) the model performance improves and provides much better forecasting properties.

Next, we investigate whether differences in forecasting performance among different models are statistically significant. For the best forecasting model (TS2), we can reject the test of equal prediction errors only for 5 (DM-Test) to 4 (GW-Test) competitive models. The random walk model (RW), the term structure model with activity indicators (TS4) and the pure survey measure (ZEW1) are rejected by both test types. Using the DM test, also the term structure models TS1 and TS3 display significant forecasting performance. The GW-test indicates significant differences relative to the PC1 model. Comparing the TS2 model with the AR benchmark model, we cannot reject the null of equal forecasting accuracy.

When we take the best survey model (ZEW3) as our reference model, we can reject the hypothesis of equal forecasting performance relative to 4 alternative models: ZEW1, TS4, RW and the AR model. However there is little evidence that the ZEW3 model does significantly better than any of the Phillips curve models or the remaining term structure models. But note that in contrast to the TS2 model the ZEW3 survey model can beat the AR model.

From table 3 further conclusions can be drawn from other inflation models. First, in contrast to Atkeson and Ohanian (2001) the random walk (RW) model performs badly compared to other models (at least for this sample period). Second, all Phillips curve models display lower RMSEs than the benchmark AR models. The differences among different specifications (unemployment rates or production) are small. The Phillips curve model with detrended production as well as changes

[14] In this case (with $R/P = 1$), West's (2006) recommendation is to divide the test statistic by a factor $\lambda = \sqrt{2/3}$.

Table 3: Forecasting performance of alternative models

		P-Values				
		Ref: TS2		Ref: ZEW3		Ref: ZEW3
Model	RMSE	DM-Test	GW-Test	DM-Test	GW-Test	ENC-Test
AR	0.9768	0.22	0.24	$0.00^{***,\dagger}$	0.03^{**}	0.96
RW	1.1824	0.01^{***}	0.01^{***}	$0.00^{***,\dagger}$	0.00^{***}	0.70
PC1	0.9450	0.33	0.01^{***}	0.77	0.64	0.29
PC2	0.9336	0.39	0.38	0.86	0.65	0.38
PC3	0.9323	0.11	0.76	0.87	0.69	0.65
TS1	0.9657	0.10^{*}	0.39	0.62	0.99	0.10^{*}
TS2	0.8922			0.73	0.73	0.10^{*}
TS3	0.9857	0.08^{**}	0.36	0.44	0.94	0.22
TS4	1.1298	0.01^{***}	0.06^{**}	0.04^{**}	0.02^{**}	0.64
ZEW1	1.2913	0.01^{***}	0.06^{**}	0.01^{***}	0.01^{***}	0.69
ZEW2	0.9823	0.29	0.43	0.14	0.49	–
ZEW3	0.9194	0.73	0.73	–	–	–

Notes: All results refer to simulated out-of-sample forecasts as discussed in the text. The different model specifications can be found in table 6. Column 2 corresponds to the annualized root mean square error (RMSE) computed for each model. Columns 3-7 contain p-values obtained from different pairwise tests for varying reference forecasts. DM-Test, GW-Test and ENC-Test refers to Diebold-Mariano-Test, Giacomini-White-Test and Encompassing-Test, respectively. † indicates the Clark and West adjustment for nested models. ***, ** and * correspond to the 1%, 5% and 10% significance level, respectively. Out-of-sample period: 2000.9 to 2008.7.

in production obtains the lowest RMSE. Third, those term structure models provide a better fit that only contain one indicator. By combining Phillips curve specifications with term structure information the RMSE increases (possibly due to overfitting).

To check robustness of our result we conduct some simple stability tests. Due to certain unique events one forecasting model may display lower and significant forecasting errors than others, but later in time the model performs worse. So we ask whether our presented results remain stable over time. Therefore we split our out-of-sample period in the middle to investigate whether test results change significantly. Table 7 reports the results of the tests for different models. Generally, we find no indication that dramatic changes occur for the considered models AR, ZEW3 and TS2 during the out-of-sample period.

Table 3 (last column) also provides some evidence whether other models contain useful information not included in survey expectations and lagged inflation. This encompassing tests indicate that Phillips curve models and univariate time series models do not provide additional information. Instead, term structure models (TS1 and TS2) may provide some information not included in the survey based inflation measure (at least when one considers a 10% significance level). These findings imply that economic experts who partizipate in the ZEW survey use methods similar to autoregressive models and Phillips curve models to form their inflation expectations. But they do not fully take into account information from interest rates or term spreads.

An obvious extension would be to combine information of survey data with those of the term structure. Since augmenting existing models with many additional regressors most likely result in overfitting and poor forecasting performance other options may be employed. In order to avoid these problems one can use methods which combine important information from many variables. One possibility is model combination (e.g. Wright, 2008) and another is factor analysis (e.g. Stock and Watson, 1999). Since survey expectations can be seen as an information averaging itself, it may be advantageous to combine it with information not included in the series, namely financial indicators such as the term spread.

5 Conclusion

This study shows how a monthly indicator for German inflation expectations with a fixed horizon may be obtained. We show basic properties of this series by applying tests of rationality as well as of Granger causality. While the concept of rationality is empirically rejected, the Granger test indicates that the indicator may be useful for forecasting inflation. An out-of-sample experiment is conducted to compare forecasts based on survey measures with other standard inflation models. Forecasts based on raw survey expectations perform poorly compared with other models, but once these survey measures are combined with lagged actual inflation rates this specifications beats almost all other specifications in terms of RMSE. Statistically, we can only differentiate among some of the models. Further tests of forecasting encompassing reveal that survey measures already contain information of most of the models, e.g. Phillips curve specifications. But there is some indication that survey data do not fully include information on financial variables. This suggests that the forecasting performance could be further improved by combining models with survey indicators with those of financial variables such as interest rates and term spreads.

However this study is limited in its validity concerning certain aspects. First, the analysis is based on CPI inflation only. So we do not provide evidence whether our survey indicator reveals information about other inflation measures such as the GDP deflator or the deflator for private consumption (since these indicators are not available for quarterly frequency). However other choices are available at monthly frequency such as the HICP definition or some subindicators like "core inflation". Second, we exclusively focus on a six month forecasting horizon since our measure for inflation expectations is exactly defined this way. Whether this survey measure might be also useful for other time horizons has to be further examined. Third, we restrict the analysis to simple single equation models. We do not consider methods of forecasting combinations or factor models. Further, multivariate models – such as VARs – are also excluded because often they were not found to provide any improvement compared to univariate time series models. Fourth, our results only apply to the specific period after reunification and the out-of-sample period 2000 to 2008. So the stability of this models is not guaranteed in the future.

In spite of all restrictions, this study indicates that quantitative inflation expectations might be a useful indicator for future inflation. Whether the results hold for other surveys as well (e.g. the consensus economics forecasts) has to be subject to future work.

References

Anderson, O. (1952). The business test of the ifo-institute for economic research, munich, and its theoretical model, *Revue de l'Institut International de Statistique* **20**: 1–17.

Ang, A., Bekaert, G. and Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better?, *Journal of Monetary Economics* **54**: 1163–1212.

Atkeson, A. and Ohanian, L. E. (2001). Are Phillips Curves Useful for Forecasting Inflation?, *Ferderal Reserve Bank of Minneapolis Quarterly Review* **25**: 2–11.

Berk, J. M. (1999). Measuring inflation expectations: A survey data approach, *Applied Economics* **31**: 1467–1480.

Carlson, J. A. and Parkin, M. J. (1975). Inflation expectations, *Economica* **42**: 123–138.

Clark, T. E. and West, K. D. (2006). Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis, *Journal of Econometrics* **135**: 155–186.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models,, *Journal of Econometrics* **138**: 291–311.

Diebold, F. X. and Mariano, R. S. (1996). Comparing predictive accuracy, *Journal of Business and Economic Statistics* **78**: 67–77.

Fama, E. F. (1975). Short-term interest rates as predictors of inflation, *American Economic Review* **65**: 269–282.

Fisher, J. D. M., Liu, C. T. and Zhou, R. (2002). When Can We Forecast Inflation?, *Ferderal Reserve Bank Chicago Economic Perspectives* **(Q1)**: 30–42.

Forsells, M. and Kenny, G. (2004). Survey expectations, rationality and the dynamics of euro area inflation, *Journal of Business Cycle Measurement and Analysis* **1**: 13–41.

Franz, W. (2005). Will the (German) NAIRU Please Stand Up?, *German Economic Review* **6**: 131–153.

Galí, J. and Gertler, M. (1999). Inflation dynamics: A structural econometric analysis, *Journal of Monetary Economics* **44**(2): 195–222.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability, *Econometrica* **74**: 1545–1578.

Harvey, D. I., Leybourne, S. J. and Newbold, P. (1998). Tests for Forecast Encompassing, *Journal of Business and Economic Statistics* **16**: 254–259.

Heinemann, F. and Ullrich, K. (2006). The impact of EMU on inflation expectations, *Open Economies Review* **17**: 175–195.

Kilian, L., Inoue, A. and Kiraz, F. B. (2006). Do actions speak louder than words? household expectations of inflation based on micro consumption data, *Discussion Paper Series 2006,26*, Deutsche Bundesbank.

Kozicki, S. (1997). Predicting real growth and inflation with the yield spread, *Federal Bank of Kansas City Economic Review* **82**: 39–57.

Mestre, R. (2007). Are survey-based inflation expectations in the euro area informative?, *Working Paper 163*, European Central Bank.

Mishkin, F. S. (1990). What does the term structure tell us about future inflation, *Journal of Monetary Economics* **25**: 77–95.

Pesaran, M. H. (1985). Formation of inflation expectations in British manufactering industries, *Economic Journal* **95**: 948–475.

Pesaran, M. H. (1987). *The Limits to Rational Expectations*, Basil Blackwell.

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation, *Jounral of Monetary Economics* **44**: 293–335.

Theil, H. (1952). On the time shape of economic microvariables and the munich business test, *Revue de l'Institut International de Statistique* **20**: 105–120.

West, K. D. (2006). Forecast evaluation, *in* G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Forecasting*, Vol. 1, Elsevier, Amsterdam, chapter 3, pp. 99–134.

Wright, J. H. (2008). Forecasting US inflation by Bayesian model averaging, *Journal of Forecasting* **forthcoming**.

A Appendix

A.1 Unit root tests

Table 4: Unit root tests for inflation

Variable	ADF-Test	PP-Test	KPSS-Test
$\Delta\pi_t$	-14.64^{***}	-14.65^{***}	0.58^{**}
$\Delta E_{t-6}\pi_t$	-14.45^{***}	-14.62^{***}	0.48^{**}
π_t	-3.82^{***}	-3.86^{***}	0.66^{**}
$E_{t-6}\pi_t$	-4.42^{***}	-4.50^{***}	0.59^{**}

Each test does include a constant term. The Lag length of the ADF-test is 0 (based on SIC). The PP test is calculated with Newey and West standard errors using a Bartlett kernel. ***: 1% and **: 5% significance level

A.2 Static Test of Unbiasedness

We also report a static test of rationality that is based on the unbiasedness assumption. As noted in the main text this type of test is only meaningful in the absence of measurement errors. If this assumption holds the test involves regressing the inflation series on the series on inflation expectations and testing the joint hypothesis that the constant is equal to zero and the coefficient of the expectations variable is equal to one (Mestre, 2007). More exactly we estimate the equation

$$\pi_t = \alpha_1 + \beta_1 E_{t-6}\pi_t + u_t. \tag{9}$$

This equation constitutes one of the earliest tests for rationality. A related test (Mestre, 2007) checks whether a model of inflation is actually able to provide good forecasts. This is what he describes with "the forecastability of inflation" and rests on the equation

$$E_{t-6}\pi_t = \alpha_2 + \beta_2 \pi_t + v_t. \tag{10}$$

Whenever the forecast error is too large, the forecast gets useless. This is tested by the null hypothesis that $\alpha_2 = 0$ and $\beta_2 = 1$.

Table 5 reports the results of both types of tests. The first one is rejected (at least at the 5% level), so one may conclude that that expectations are indeed biased, although they are constructed that the changes of inflation expectations are on average unbiased (see section 2). The second test does not reject the null, indicating that forecastability of inflation is given.

Table 5: Static test for Rationality

Regression	Coefficients α_i	β_i	F-Test $\alpha = 0$, $\beta = 1$	p-value
(9)	0.39	0.77	4.52	0.01
(10)	0.40	0.82	2.15	0.12

The tests are based upon Newey-West standard errors.

A.3 Models and stability test

Table 6: Model specifications

	Abbr.	Specification (included regressors)
Univariate Time Series Models	AR	π_t^6, π_{t-6}^6
	RW	$1/2\left(\pi_t^6 + \pi_{t-6}^6\right)$
Phillips Curve Models	PC1	π_{t-6}^6, U_{t-2}, *seasonal dummies*
	PC2	π_{t-6}^6, $Ugap_{t-1}$, *seasonal dummies*
	PC3	π_{t-6}^6, $Ogap_{t-1}$, $\Delta iprod_{t-1}$, *seasonal dummies*
Term Structure Models	TS1	π_t^6, r_{t-4}^s, *seasonal dummies*
	TS2	π_{t-6}^6, $(r^l - r^s)_{t-3}$, *seasonal dummies*
	TS3	π_t^6, r_{t-4}^s, $(r^l - r^s)_{t-3}$, *seasonal dummies*
	TS4	π_t^6, π_{t-6}^6, U_{t-2}, U_{t-6}, r_{t-1}^s, r_{t-5}^s, $(r^l - r^s)_{t-1}$, $(r^l - r^s)_{t-6}$, *seasonal dummies*
ZEW inflation expectation	ZEW1	$E_t\pi_{t+6}^6$ (raw data, without estimation)
	ZEW2	$E_t\pi_{t+6}^6$, *constant*
	ZEW3	π_t^6, π_{t-6}^6, $E_t\pi_{t+6}^6$

Table 7: Break test

Model	P-Value
ZEW3 - AR	0.50
TS2 - AR	0.70
ZEW3 - TS2	0.52

A break point in the middle of the sample is tested by using the DM-Test from Table 3 augmented by a dummy variable that is equal to 1 before the break point and 0 afterwards. The P-Value corresponds to the t-test obtained for the dummy variable computed with HAC standard errors.

A.4 Data description

Most series were taken from the Deutsche Bandesbank database. The following abbreviations are used in the data description: SA = seasonally adjusted; NSA = not seasonally adjusted; HWWI = Hamburgisches WeltWirtschafts Institut (additional data source); ZEW = Center for European Economic Research (additional data source)

Definitions

P	consumer price index: total index (2005=100, NSA)
U	unemployment rate (SA)
$Ugap$	HP(14400)-filtered unemployment rate using only past information (SA)
$iprod$	industial production: manufactering (2000 = 100, SA)
$Ogap$	HP(14400)-filtered industrial production using only past information (SA)
r^l	long term government bond yield: 9-10 years maturity (NSA)
r^s	money market rates reported by Frankfurt banks: Three-month funds (NSA)
P^{raw}	HWWI commodity price index for Euro area (euro basis, NSA)
A	fraction of respondes reporting "inflation goes up" (ZEW, NSA)
B	fraction of respondes reporting "inflation goes down" (ZEW, NSA)