Surveys and Econometricians
When can surveys beat econometricians in forecasting inflation?

Krisztina Molnar* and Zoltan Reppa†
Work in Progress
April 21, 2009

Abstract
We examine the forecasting power of surveys in different environments, countries with stable inflation and countries where inflation is more volatile. We test whether surveys forecast inflation better than the best out of sample VAR forecaster. We find a very clear pattern: surveys are better real time forecasters than a VAR in stable economies, and not in countries with a highly volatile inflation.

JEL classification: C62, D83, D84, E52
Keywords: Rational Expectations, Survey expectations

Introduction
Rational expectations paradigm has gained a substantial prominence in economics, yet empirical evidence is mixed, and testing rational expectations is still a matter of discussion. "A general point is that it is difficult to say anything definite about whether market participants have formed expectations rationally without a clear understanding of the process determining" movements in the underlying process. (Cook and Hahn (1990) p.14)

However, once we depart from fully rational expectations, there are many ways to do so. In this paper we propose that one way to "tie" our modelling hands is

---

*Norwegian School of Economics and Business Administration, Bergen, Norway; Email: krisztina.molnar@nhh.no
†Magyar Nemzeti Bank, Budapest, Hungary; Email: reppaz@mnb.hu.
to examine whether there are some clear patterns in how people form their expectations in different environments. We examine survey expectations of inflation in different countries, and compare them with econometric forecasting models.

Since there is no clear understanding about the underlying model of inflation, we think it is important to evaluate survey expectations in real time. The basic idea of rational expectations is that agents should be forward looking and should use all available information in an optimal way. We propose to mimic this mental process in real time. All available information can include quantitative and non-quantitative information too. In this paper we restrict our attention to quantitative information. We assume the set of all available information is past data, we construct a good econometrician who combines this dataset to produce out of sample forecasts of inflation and compare survey forecasts to this good econometrician.

Our methodology is to find the best out of sample forecaster within the set of Vector Autoregression models. We also want to make sure, that we find forecasting methods that are good compared to other sets of econometric models. Therefore, we compare forecasting performance to the best forecasting models of inflation found in the latest literature (Ang et al. (2007)).

Some papers have already documented that US surveys outperform econometric models (Stock and Watson (1999), Ang et al. (2007)). The paper of Ang et al. (2007) examines several survey measures of inflation for the US. They compare the forecasting power of surveys to a wide set of econometric models. Out of sample forecasts are generated with time series ARIMA models, regressions using real activity measures motivated by the Phillips curve, and term structure models.

We reinforce their finding on US data, and in addition we show that this finding is not a general feature of survey expectations. In some countries surveys outperform econometric models, in other countries they do not. The interesting finding is that there is a clear pattern. Surveys outperform the best econometric forecasts only in stable economies. In countries where the volatility of inflation is high, or there were several structural breaks during our sample period, surveys do worse then econometricians. Surveys are real time rational only in stable economies.

Our paper also questions whether conventional econometric tests that reject rationality in inflation expectations constitute evidence against using all available information in an efficient way.

Several authors emphasize that traditional tests of biasedness in inflation expectations, testing whether forecast errors are zero on average, can be misleading. Partly because the results depend very much on the sample size, partly because it is well known that sluggish expectation formation can be optimal in environments which feature some type of information friction (see for example Muth (1960), Webb (1987), Andolfatto et al. (2005)).

We examine pitfalls in efficiency tests. We use conventional tests of rationality
to test the out of sample forecasts of the best econometric predictor. The interesting finding is that econometric predictors can be inefficient even with respect to macro variables that were indeed used in the estimation. The intuition for this is, that the coefficients of the econometric model are reestimated in each period. Because of the changing coefficients, the out of sample forecasts are in general not orthogonal to the information used in the regression. If for example if there is a structural break in the underlying data generating process, the estimated coefficients will be drifting. After a structural break it takes time to learn the new coefficients from the real time data. During this learning period the real time forecast errors will reflect some influence of the underlying data, even if this data is used in the econometric forecast.

We find that a further pitfall in conventional econometric tests of efficiency. If a forecast is inefficient with respect to a macro variable, it does not provide evidence that the forecasts could have been improved by using this variable.

The paper is organized as follows. In section 1 we give an overview of the empirical research on rationality in survey expectations. Section 2 discusses out methodology, section 3 describes the dataset. Section 4 shows estimation results. First we examine US survey data, then we turn to a set of wider countries.

1 Literature on Testing Rational Expectations

In this section we discuss conventional tests of rationality. We also discuss the empirical research on rationality in survey expectations of inflation.

Assuming all information is available and transparent: the economic model and the type of shocks are known, rational expectations correspond to the true mathematical conditional expectation, conditional on the information set. In mathematical notation we have

\[ \pi_{t+1}^r = E_t(\pi_{t+1}|I_t) \]  
\[ \pi_{t+1}^r = \pi_{t+1} + e_{t+1}, \]  

where \( \pi_t \) is inflation at time \( t \). Expectations are denoted with subscript \( e \), and \( I_t \) is the information set at time \( t \). This is sometimes referred to as a perfect foresight with error expectations model. Testing rationality then can be done in the following way:

\[ E_t(e_{t+1}) = 0 \quad \text{no bias} \]  
\[ E_t(e_{t+1}|I_t) = 0 \quad \text{no inefficiency of information use} \]
The information set includes all the variables that would be contained in a sophisticated model of expectation formation.\(^1\) If some variable in the information set significantly correlate with the forecast errors, test (4) suggests that agents did not take into account all the relevant information in this variable in producing their forecasts.

Muth (1961) proposed that “expectations, since they are informed predictions of future events, are essentially the same as the predictions of the relevant economic theory”. Muth did not define rationality as good ex post forecasting accuracy. Rationality implies good forecasts conditional on the information set. The conventional tests, assume that the underlying model is part of the information set, therefore we can test rationality as (3) and (4).

In reality the true data generating process is not in the information set. Conventional tests, (3) and (4) are therefore not the perfect tests, but one can still argue to test rationality by examining forecast accuracy in the following sense. People aim to forecast what will occur, and therefore have strong incentive to use rules that work well. In recurrent situations, people should understand how the future unfolds from the past and adjust their expectations (Sargent (1993)). Therefore rational forecasts should not deviate from the ex post outcome systematically, in a predictable way. Our paper follows the recommendation of Sargent and examine forecast accuracy conditional on past patterns in the data.

Several papers have tested rationality with the above model.\(^2\) In early studies, the existence of strong serial correlation in the forecast errors of survey measures of inflation was commonly interpreted as a sign of irrationality. However because of overlapping periods, survey respondents are unaware of their forecast errors for a time, and this in itself causes autocorrelated errors. Later studies have corrected for serial correlation in the error terms to perform statistical tests, and more studies have found rationality in surveys.\(^3\)

Bias tests are typically sensitive to the sample period. For example Thomas

---

\(^1\) A distinction is commonly made between weak- and strong-form of efficiency. Weak-form of efficiency is based on the notion that past values of \(p\) are costless, while other variables are costly. Expectations are weak-form efficient if they incorporate all the information in the history of \(p\). Strong-form of efficiency indicates that agents have effectively incorporated all variables that economic models of \(p\) would include. Some economists define efficiency differently. Weak-form of efficiency as taking account of all publicly available information and strong-form efficiency as taking account of additional, more sophisticated information, for example forecasts generated by a model.


\(^3\) For a study testing rationality with cointegrating methods see for example Grant and Thomas (1999).
(1999) and Mehra (2002) report that for subperiods\textsuperscript{4} unbiasedness can be rejected, while for the whole sample not. The reason for this is that surveys tend to over-predict inflation when inflation is falling, and underpredict inflation when inflation is increasing.

These papers also report that efficiency tests are similarly sensitive to subperiods. Thomas (1999) for example rejects efficiency for the whole sample period\textsuperscript{5}, but after the 1980’s he finds some signs that agents improved the utilization of information. Several other papers reject efficiency of survey expectations. For example Baghestani (1992), Ball and Croushore (1995) show that survey forecasts do not make efficient use of all available information.

2 Methodology

In this section we describe how do we construct the “best” econometric forecasting algorithms. Our aim is to find a predictor that provides the best real time forecasts and then compare its forecasting power to surveys.

We would like to construct an econometric model that utilizes all the available information and provides the best out of sample forecasts on average. Of course the set of econometric models is almost uncountable, therefore we have to restrict our analysis to a subset of them.

To make sure our best econometric model has good forecast accuracy, we rely on the existing literature on inflation forecasting. A recent paper Ang et al. (2007) have examined a wide set of econometric models, including term structure models, time series ARIMA models, and regressions using real activity measures. They have found that ARMA and Random Walk time series models provide the best out of sample forecasts of inflation. Therefore as a benchmark we will always compare our results to the out of sample forecasts of Random Walk and ARMA. We have experimented with ARMA models up to 6 lags in both in the AR and MA part. We report only the specification with the best forecasting performance.

We search for the the best real time predictor on the set the set of Vector Auto Regressions with discounting past data. The equations of the VAR-s were estimated by weighted least squares with exponentially declining weights; i.e. they


\textsuperscript{5}For the whole sample period weak-form of efficiency is not rejected, but strong form of efficiency is rejected, because agents do not use information contained in the output gap.
minimize
\[ \sum_{t=1}^{T} \alpha^{T-t} (y_{t}^{i} - x_{t}\beta)^{2} = \sum_{t=1}^{T} (\alpha^{(T-t)/2}y_{t}^{i} - \alpha^{(T-t)/2}x_{t}\beta)^{2}, \] with \(0 < \alpha \leq 1\). \hspace{1cm} (5)

Here \(y^{i}\) is the \(i^{th}\) variable in the VAR and \(x\) contains the lags of the variables.

We determine \(\alpha\) by a grid search method, which involves the following steps:

1. First we choose the time interval \([T_{1}, T_{2}]\) that is used to find the best fitting VAR forecasts. (Equal to the survey sample.)
2. Choose a VAR. In other words choose what variables are included in the VAR.
3. Fix a discounting parameter \(\alpha\).
4. Calculate one-year-ahead forecast of inflation with this VAR and \(\alpha\).
5. Calculate root mean square deviation (RMSD) between the VAR forecasts and the actual inflation data.
6. We repeat steps 2 - 5 for different values of \(\alpha\) and different VARs, and chose the one with the minimal RMSD.

We do the estimations in a recursive fashion (see Appendix for details). The initial values we estimate from the data before \(T_{1}\). For example if the sample of the survey of inflation expectations is from \(T_{1}\) to \(T_{2}\), we are searching for the best econometric forecast on this sample. We set the initial values for the recursive forecast estimates from the data before \(T_{1}\). For each VAR and each value of discounting past data we estimate the VAR coefficients on the data before \(T_{1}\). After this we estimate the new coefficients in a recursive fashion.

We examine VARs on a wide set of macro variables. We search for the best performing VAR, both by optimizing on what variables should be included in the VAR and both by optimizing how much past data should be discounted. Therefore we find the VAR that would have produced the forecasts closest to perfect foresight on a given sample.

We always find the best VAR on the same sample as the surveys, in order to be able to compare their forecasting performance. This is important since in different periods forecasting inflation might have been easier then in other periods.
3 Data

This section describes the dataset.

We use a wide variety of macro variables (see below). We include all macro variables that are typically used in VARs to forecast inflation. Further we include all macro variables that are found to be good in explaining forecast errors of surveys.

- **Economic data for the US** CPI, Industrial production, Unemployment rate, Interest rates (Federal funds effective rate, 3 and 6 month TB, 3 and 5 year TY), Money supply(m1, m2), Wage (average hourly real earnings), Oil price (PPI, fuels and related prod)

- **Surveys US** Michigan Survey, Survey of Professional Forecasters

- **Other surveys** Australia, Israel, Czech Republic, Hungary, New Zealand, Poland

We examine several countries besides the US: Australia, Czech Republic, Hungary, Israel, New Zealand, and Poland. We use the same set of variables in these estimations as for the US estimates, whenever they are available. We always included the short term interest rate that is controlled by the central bank. In addition we tried to include short term and long term market interest rates too. For details we refer the reader to the Appendix.

If available we use monthly data instead of quarterly for the estimations. The motivation for this is that a good forecaster should use all available data. Even if the data for the last quarter is not published yet, data on some months might be already available.\(^6\) As a robustness check we repeat the analysis on quarterly data. For Australia and New Zealand only quarterly CPI data is available, therefore we ran only quarterly regressions. For Israel we are also running only quarterly regressions.\(^7\)

Bellow we describe the surveys we used in our analysis. Czech survey data is from a survey conducted by the Czech National Bank. Hungarian survey expectations are forecasts of professional forecasters from the Reuters survey. For Israel we use survey data of inflationary expectations conducted by the Bank of Israel. The Polish survey is conducted by the National Bank of Poland. The New Zealand survey is conducted by the Reserve Bank of New Zealand. For the US we use the Survey of Professional Forecasters and the Michigan survey.

For Australia we analyze the Melbourne Institute Survey of Consumer Inflationary Expectations. It is a random sample of 1200-1400 households, respondents

---

\(^6\)The Ang et al. (2007) was conducted on quarterly data.

\(^7\)We are currently working on the monthly regressions.
are asked a range of questions. The Melbourne Institute makes some adjustments to the raw data in estimating the median value of expected inflation. They convert qualitative responses to quantitative ones. People are asked “By this time of the year, do you think the prices of things you buy will go up or down? If up, by how much? If down, by how much?” About 10-20 percent of the respondents give only qualitative answers, these are converted into quantitative answers by reallocating them using the distributions of quantitative responses. The second adjustment is to correct for sampling bias, according to gender, age and location. Median one year ahead inflation expectations are available on a quarterly basis from the website of the Reserve bank of Australia from June 1993.

The Czech National Bank (CNB) started regular measurement of inflation expectations among financial market analysts in May 1999. In June 1999 the survey was extended to businesses and households. Respondents are asked about their expectation of the year-on-year CPI for one- and three-year horizon. The financial survey is conducted monthly among 15 selected domestic and foreign analysts, who trade both on the money and on the capital markets. Our sample period is monthly from May 1999 to December 2008. The survey of businesses and the survey of households are performed quarterly. Surveys are conducted each March, June, September, and December. The business survey covers 120 managers of major businesses from all sectors of the economy. Our sample is June 1999 to December 2008. The quarterly surveys of households are based on a large and randomly selected sample of 600 households. Our sample period is June 1999 to March 2007 with quarterly frequency. Starting in July 2007, the CNB abandoned its quantitative approach to monitoring inflation expected by households and started using a qualitative assessment. For all three surveys medium responses are available on the website of the Czech National Bank.

For Hungary we use the Reuters survey. The Reuters survey asks about end of year forecasts, and average yearly forecasts. Expectations of one year ahead inflation were calculated by the National Bank of Hungary with a projection method. Data are available from the first quarter of 1997 until the second quarter of 2005.

---

8The number of financial market analysts surveyed has been gradually increased from 10 to 15. The forms are sent to the CNB on the tenth working day of the current month.

9From July 2007 Households are asked qualitative questions about their perception of past inflation and their expectation of future inflation. These indicators are collected as part of the European Commission Business and Consumer Survey.

10Up until March 2003, outliers in the responses of businesses and households were not excluded; expectations were computed as the arithmetic mean of all the answers. Since June 2003, 5% of the maximum responses and 5% of the minimum responses of businesses and households have been removed and average expectations have been computed as the arithmetic mean of the remaining 90%.

11We would like to give thanks for Balazs Vonnak for providing us with the data.
Data after 2005 is less reliable.\textsuperscript{12}

Israel Companies Survey The Companies Survey is a quarterly survey of about 700 companies. Participants are asked to report on the direction of changes in different variables (increase, decrease, or no change), and on the degree of change (great, slight). We use mean figures calculated by the Bank of Israel.\textsuperscript{13} The survey’s sample period is from the first quarter of 1997, to the 3rd quarter of 2008. Survey answers were collected in each March, June, September and December.

The Reserve Bank of New Zealand Survey of Expectations is a New Zealand-wide quarterly survey of business managers. There are approximately 200 respondents. They are asked about their forecast on several macroeconomic variables, among them CPI inflation on several horizons. We use expectations of the average one-year-ahead (CPI) inflation. The Reserve Bank of New Zealand reports the mean expectation levels of respondents. Our sample period is 3q1987-4q2008. Surveys are conducted in the middle of each quarter.

We downloaded Polish inflation expectations from the website of the National Bank of Poland. These are consumer inflation expectations from the Ipsos-Demoskop survey. The survey is carried out monthly, on a sample of approximately 1,000 individuals. The survey questions concerning inflation expectations are qualitative. Quantitative forecasts are constructed by the National Bank of Poland. For details see Lyziak (2003).

For the US we use the Michigan Survey and the Survey of Professional Forecasters. The Michigan Survey of Consumer Attitudes and Behavior surveys a cross-section of the population on their expectations over the next year. The Michigan survey is conducted monthly from January 1978. On average 500-700 consumers are surveyed each month. Survey of Professional Forecasters (SPF) covers more sophisticated analysts. Survey respondents are economists working in industry and professional forecasters. This survey is conducted quarterly from 3Q 1981. Survey responses were collected in the middle of the second month in each quarter.

4 Results

In this section we show estimation results. We compare surveys to the the best VAR forecaster, where the best forecaster is always searched on the same sample length as the sample of the survey. First we examine results for the US. Next we discuss results for a wider set of countries.

\textsuperscript{12}In 2005 there was a tax reform in Hungary. Estimates of inflation expectations depend on assumptions on how this tax reform is incorporated into inflation expectations.

\textsuperscript{13}We would like to give thanks to Alex Ilek for providing us the data.
4.1 Results on US data

For the US we have examined two surveys, the Michigan and the Survey of professional forecasters. There have been an extensive research on these surveys, to which we can compare our results.

First we show how the forecasting performance of the best VARs compare to surveys, then we turn to examining bias and efficiency of the forecasts.

The two boxes below show the best VAR forecasting specifications on the Michigan and SPF sample. Ang et al. (2007) have examined time-series ARIMA models, regressions using real activity measures motivated from the Phillips curve, and term structure models that include linear, non-linear, and arbitrage-free specifications. They have shown that ARMA models are the best forecasters of CPI inflation. Therefore as a benchmark we have chosen to compare forecasting performance to ARMA and Random Walk (RW) models. We have always found the best ARMA specification on the same sample as the surveys.

<table>
<thead>
<tr>
<th>Results on the sample of the Michigan Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD of survey expectations: 1.5408</td>
</tr>
<tr>
<td>RMSD of Random Walk: 1.8904</td>
</tr>
<tr>
<td>RMSD of ARMA(4,4) forecast error: 1.7172</td>
</tr>
<tr>
<td>RMSD of best VAR: 1.6705</td>
</tr>
<tr>
<td>Best VAR: dlog(cpi) dlog(ind), with 10 lag, optimal $\alpha$ 0.9827</td>
</tr>
</tbody>
</table>

We have found that the best performing VARs outperform ARMA and RW models. ARMA and RW models were found best forecasters in Ang et al. (2007), our best VARs outperform these because we are optimizing on the set of VARs. For a given sample we are searching for the best out of sample forecaster VAR algorithm; we optimize both in terms of what variables should enter into the VAR and how much past data should be discounted. In contrast Ang et al. (2007) estimate a wide set of models, including one type of VAR without allowing for discounting of past data. Since our best VARs outperform the best forecaster found in Ang et al. (2007), we contrast forecasting performance of surveys with an even stricter requirement then them. In other words we are biasing the forecasting competition between surveys and econometric models against surveys.

<table>
<thead>
<tr>
<th>Results on the sample of the Survey of Professional Forecasters</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD of survey expectations: 1.2279</td>
</tr>
<tr>
<td>RMSD of Random Walk: 1.5800</td>
</tr>
<tr>
<td>RMSD of ARMA(4,6) forecast error: 1.3877</td>
</tr>
<tr>
<td>RMSD of best VAR: 1.2420</td>
</tr>
<tr>
<td>Best VAR: dlog(cpi) dlog(ind) w, with 8 lag, optimal $\alpha$ 0.9815</td>
</tr>
</tbody>
</table>
The main thing to take away from our results is that both the Michigan and the Survey of Professional Forecasters outperform the best econometric forecasts. The Michigan surveys root mean square error is 1.54\%, while the VAR that provides the best real time forecasts on this sample has a root mean square error of 1.67\%. The difference for the professional survey is less pronounced. The best VAR on the SPF sample has a root mean square error of 1.24\%, while the SPF has a root mean square error of 1.22\%.

These results reinforce the results of Ang et al. (2007). They have shown that for a wide variety of econometric models surveys do better then out of sample econometric forecasts. Our result is for US data surveys beat VARs even if we contrast them with the VARs that would get the closest possible to the perfect foresight forecasts.

A further difference between Ang et al. (2007) and our estimations is that we use monthly data, while they use quarterly data. This again biases the results in favor of econometric models, since there more information can be gained from monthly data. We have repeated the exercise also for quarterly data, and the main result remains the same: the Michigan Survey and the SPF outperforms the best VARs. (See results in the next subsection.)

The VAR that provided the best forecasts on the Michigan sample uses CPI inflation and growth of industrial production, with a lag of 10 month and a very small discounting of past data. A high $\alpha$ indicates that on average on the it is not optimal to discount past data very much. The best VAR on the SPF sample includes CPI inflation, growth of industrial production, and wage (average hourly real earnings), with variables up to 8 month lag, and again with little discounting of past data. High level of discounting of past data is be beneficial if there are structural breaks in the data.\footnote{The use of constant gain algorithms to track structural changes is well known from the engineering literature (see Benveniste and P. (1990), Part I. Chapters 1. and 4).} For the US on the sample 1978-2008 or 1981-2008 this is not the case, the best forecasters on average should not discount past data very much.

4.1.1 Basic statistics of forecasting performance

In this section we compare in more detail the forecast errors of the best VARs and US surveys. We have already seen that on average surveys outperform the best VAR algorithm, in this section we examine the bias in forecasts and the standard deviation of errors.

Figures 4.1.1 and 4.1.1 show the best VAR forecasts together with CPI inflation. CPI inflation is shifted back with one year, therefore the difference between the two lines are actual forecast errors.
Our first observation is that VAR forecasts resemble surveys in the sense that they are lagging behind actual inflation. They tend to underpredict changes in inflation.

There have been an extensive research testing bias in survey forecasts by examining whether their forecast error is significantly different from 0. Table 1 confirms what the literature consistently reports, there seem to be a bias in survey expectations. On average the Survey of Professional Forecasters overpredicts inflation, while the Michigan Survey underpredicts it. When we compare the two surveys on
the same sample (Table 1 lower half), we can see that the opposite bias in surveys is simply due to the different sample period. After 1981 the Michigan Survey also overpredicts inflation. Interesting to note that on the same sample the median Michigan Survey has a much smaller bias than the SPF.

After 1981 US inflation have decreased substantially, and surveys overpredicted inflation. This corresponds to the stylized fact that during rising inflation surveys tend to underpredict inflation, while during periods of decreasing inflation surveys
Table 1: Basic statistics for VAR and survey forecasts

<table>
<thead>
<tr>
<th></th>
<th>VAR -Michigan</th>
<th>VAR -SPF</th>
<th>Michigan</th>
<th>SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0245</td>
<td>0.1777</td>
<td>-0.3115</td>
<td>0.2719</td>
</tr>
<tr>
<td>Std.</td>
<td>1.6727</td>
<td>1.1777</td>
<td>1.5111</td>
<td>1.2031</td>
</tr>
</tbody>
</table>

On the same sample (SPF sample)

<table>
<thead>
<tr>
<th></th>
<th>VAR -Michigan</th>
<th>VAR -SPF</th>
<th>Michigan</th>
<th>SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.2032</td>
<td>0.1777</td>
<td>0.0011</td>
<td>0.2719</td>
</tr>
<tr>
<td>Std.</td>
<td>1.3698</td>
<td>1.1777</td>
<td>1.0907</td>
<td>1.2031</td>
</tr>
</tbody>
</table>

tend to overpredict inflation.\textsuperscript{15}

However surveys did not perform bad compared to the best VAR algorithms. The best real time VAR forecasts show a similar bias then the surveys, after 1981 on average they overpredicted inflation. This implies that the bias in surveys might be only a consequence of difficulty in forecasting in this period, in other words even a sophisticated econometrician would have had the same direction of bias in its survey.

We have also tested whether these biases are statistically significant from 0. After using a Newey West correction up to 12 months, the tests could not reject the null of no bias neither in the VAR forecasts neither in surveys. This is in contrast with some earlier findings, for example N. Gregory Mankiw and Wolfers (2003), therefore we tested bias in expectations on smaller samples too. We have found that the results are sensitive to the sample period.

The upper panel of Table 1 show that the errors of VAR forecasts on the Michigan sample have bigger standard deviation then the errors of the Michigan survey. This results from the high variance of the real time forecasts of this VAR. (See also figure 4.1.1). On the SPF survey, when inflation have substantially decreased and the variance of inflation have decreased too, the forecast errors of the best VAR have slightly smaller standard deviation then the forecast errors of the SPF.

We conclude that the VARs that provide the forecasts closest to perfect foresight, behave similarly to surveys. Because of their backward looking nature, they underpredict inflation when it is rising and underpredict inflation when it is decreasing. Therefore how big is the bias in forecasts depends very much on the sample. On average the forecast bias is negligible both in the case of VARs and surveys.

\textsuperscript{15}See for example Long (1997) and Dotsey and DeVaro (1995).
4.1.2 Efficiency test of US forecasts

In this section we perform standard efficiency tests of forecasts regressing the forecast error on freely available information. If forecast errors can be explained by an available macro variable, the standard conclusion from these tests is that forecasts could have been improved by using this variable.

Tables 2 and 3 report the most interesting tests, regressing the forecast errors on wage and industrial production. Recall the the best VAR on the SPF sample uses both variables, while the best VAR on the Michigan sample uses industrial production.

Table 2: Is information in macro variables fully incorporated?

\[
\pi_t - \hat{E}_{t-12} \pi_t = \beta_0 + \beta_1 w_{t-13}
\]

<table>
<thead>
<tr>
<th></th>
<th>VAR-Michigan</th>
<th>VAR-SPF</th>
<th>Michigan</th>
<th>SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_0)</td>
<td>16.025***</td>
<td>8.096*</td>
<td>19.837***</td>
<td>12.666***</td>
</tr>
<tr>
<td></td>
<td>(2.7471)</td>
<td>(1.8817)</td>
<td>(3.1077)</td>
<td>(2.8626)</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>-2.011***</td>
<td>-1.005*</td>
<td>-2.532***</td>
<td>-1.571***</td>
</tr>
<tr>
<td></td>
<td>(-2.7044)</td>
<td>(-1.8522)</td>
<td>(-3.1105)</td>
<td>(-2.7447)</td>
</tr>
<tr>
<td>Adj.(R^2)</td>
<td>0.1392</td>
<td>0.0614</td>
<td>0.2732</td>
<td>0.0918</td>
</tr>
</tbody>
</table>

Notes: ***, **, and * denote significance at 1, 5, and 10% levels.

Newey-West t-statistics in parenthesis, correcting for autocorrelation up to one year.

The most interesting result from Table 2 is that a standard efficiency test of the best VAR on the SPF sample is inefficient with respect to wages, even though these forecasts are generated by using this variable. Yet, since the forecasts are generated out of sample at every time \(t\), reestimating the VAR on all available data, it can happen the the final forecast errors remain inefficient with respect to a variable used in the regressions. This result remains significant if we regress the forecast error on longer lags of average hourly earnings. Results also remain significant if we correct for longer autorcorrelation of the residuals.

The second result that we should take home from Table 2 is that the best VAR on the Michigan sample is also inefficient with respect to average hourly earnings. This VAR does not use wages, so it is more intuitive that its forecast errors are explainable by wages, then in the case of the SPF-VAR. This result is interesting because on the 1978-2008 sample the Mich-VAR provides the best out of sample forecasts. Therefore even though standard efficiency tests would recommend to include wage into the regression, doing so would increase the forecast error. Including wages into this VAR would increase the root mean squared forecast error from 1.67 to 1.79.\(^{16}\) These results, remain significant for longer lags and for longer correction for the autocorrelation of the regression errors.

\(^{16}\)If we allow for optimizing in the discounting of past data, the optimal \(\alpha\) decreases to 0.9784.
In addition both the Michigan and the SPF survey errors are strongly significantly explained by lagged wage. Surveys are also remain significantly inefficient for longer lags of average hourly earnings and for longer correction for the autocorrelation of the regression errors.

Table 3: Is information in macro variables fully incorporated?

\[
\pi_t - E_{t-12} \pi_t = \beta_0 + \beta_1 d\log(\text{ind})_{t-13}
\]

<table>
<thead>
<tr>
<th></th>
<th>VAR-Michigan</th>
<th>VAR-SPF</th>
<th>Michigan</th>
<th>SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_0)</td>
<td>0.0508</td>
<td>0.034</td>
<td>-0.295</td>
<td>0.273</td>
</tr>
<tr>
<td>(0.2307)</td>
<td>(0.2039)</td>
<td>(-1.3826)</td>
<td>(1.0596)</td>
<td></td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>-12.436</td>
<td>-21.259*</td>
<td>-7.303</td>
<td>-0.451</td>
</tr>
<tr>
<td>(-0.6742)</td>
<td>(-1.7055)</td>
<td>(-0.4724)</td>
<td>(-0.0307)</td>
<td></td>
</tr>
<tr>
<td>Adj.(R^2)</td>
<td>-0.0005</td>
<td>0.0072</td>
<td>-0.0018</td>
<td>-0.0096</td>
</tr>
</tbody>
</table>

Notes: *, **, and *** denote significance at 1, 5, and 10% levels.

Newey-West t-statistics in parenthesis, correcting for autocorrelation up to one year.

Table 2 show similar efficiency tests on industrial production. All surveys seem to successfully incorporate information in the industrial production, also the best VAR on the Michigan sample. The only exeption is the best VAR on the SPF sample. Industrial production of the previous month is significantly explaining the forecast error of the best VAR on the SPF sample. This is also surprising, since this VAR uses industrial production up to 8 lags. Nevertheless this result is not very strong, in the sense that for longer lags of industrial production the effect becomes insignificant. We have also experienced with different length of Newey-West correction in the error terms, results are not sensitive, the t-statistics become insignificant only for very long lengths in the error correction.\(^{17}\)

We performed further efficiency tests, on other macro variables (results not reported here). Both surveys and VAR forecasts are inefficient with respect to several other variables. For example the are strongly inefficient with respect to both short term and long term interest rates.

We conclude that standard efficiency tests can be misleading. If forecast errors can be explained with an available macroeconomic variable, it does not necessarily mean that this variable was not used in the forecast, neither that the forecasts could have been improved by using this macro variable for the forecasts.

This signals that if we force the VAR estimator to suboptimally use the wage variable, it is optimal to discount past data more heavily.

\(^{17}\)Results become insignificant at 17 month correction of the Newey West residuals.
4.2 Results for a wider set of countries

In this section we examine how do survey expectations compare to the best VARs for a wider set of countries.

Tables 4 and 5 summarize our results for quarterly regressions. Table 6 and 7 summarizes our results for monthly regressions.

We compare Transition economies and Israel to the more stable countries: US, Australia and New Zealand. The US, Australia and New Zealand are relatively stable economies during our sample period, while transition economies are less stable: variance of inflation is higher, and they have experienced regime changes in the last decade (See also tables 4 and 6.). The Czech Republic is an interesting case. The volatility of inflation is relatively low, in this sense the Czech economy is more similar to the US then other transition economies. Yet also in the Czech Republic there were several regime changes in monetary policy during our sample. Israel experienced hyperinflation during our sample period.

In transition economies the economic environment is volatile from the nature of the transition itself and also because there were a regime changes of the exchange rate system. Initially transition economies had some form of fixed exchange rate and later switched to inflation targeting with free of managed float of the exchange rate.\textsuperscript{18}

The first result to notice is that there is a difference between countries with small and more volatile inflation.

US is the country with the smallest standard deviation of inflation. US surveys perform better then econometric forecasts both on a monthly and on a quarterly frequency. The SPF have smaller forecast errors then the Michigan survey on average. On a monthly frequency the best econometric forecast on the SPF sample (1981-2008) is the best ARMA forecast.

New Zealand surveys also attain a smaller forecast error then the best VAR. In New Zealand the variance of inflation on the whole sample is quite high, 5.12, yet the standard deviation of inflation on the survey expectations’ sample is only 1.6626. In Australia, the variance of inflation is a bit higher, then in the US. The variance of CPI inflation is even higher (6.16) on the survey sample. Yet, surveys still perform almost as well as the best VAR forecast.

In Hungary and Poland the variance of inflation is much higher then in the US, also there were several regime change of monetary policy during our sample. Surveys clearly do much worse them the best econometric models.

In the Czech republic the variance of inflation is lower then in other transition economies.\textsuperscript{18} Hungary experienced one big change in monetary policy during our sample period: in the 2nd quarter of 2001 Hungary abandoned the crawling peg exchange regime and switched to inflation targeting.\textsuperscript{19} The Czech National Bank switched to inflation targeting in 1998. Poland switched to inflation targeting in 1999.
Table 4: Survey errors and errors of the best VAR

<table>
<thead>
<tr>
<th></th>
<th>Survey</th>
<th>Best VAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA spf</td>
<td>1.1825</td>
<td>1.5884</td>
</tr>
<tr>
<td>USA mich</td>
<td>1.5361</td>
<td>1.5432</td>
</tr>
<tr>
<td>Australia</td>
<td>1.3256</td>
<td>1.22886</td>
</tr>
<tr>
<td>New Zealand</td>
<td>1.5286</td>
<td>1.8326</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>1.6987</td>
<td>0.5588</td>
</tr>
<tr>
<td>Hungary</td>
<td>1.9136</td>
<td>0.9772</td>
</tr>
<tr>
<td>Poland</td>
<td>2.4152</td>
<td>2.0276</td>
</tr>
<tr>
<td>Israel</td>
<td>3.5435</td>
<td>1.2993</td>
</tr>
</tbody>
</table>

Table 5: Best VARs

<table>
<thead>
<tr>
<th>Variables</th>
<th>α</th>
<th>Lags</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA spf</td>
<td>dlog(cpi)dlog(gdp)tbr</td>
<td>0.9405</td>
</tr>
<tr>
<td>USA mich</td>
<td>dlog(cpi)dlog(gdp)tbr</td>
<td>0.9868</td>
</tr>
<tr>
<td>Australia</td>
<td>log(cpi)log(gdp)log(m3)</td>
<td>0.9913</td>
</tr>
<tr>
<td>New Zealand</td>
<td>dlog(cpi)dlog(gdp)w</td>
<td>1</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>dlog(cpi)dlog(gdp)tbr</td>
<td>0.9970</td>
</tr>
<tr>
<td>Hungary</td>
<td>dlog(cpi)dlog(gdp)tbr</td>
<td>0.8589</td>
</tr>
<tr>
<td>Poland</td>
<td>log(cpi)log(gdp)^r − lendlog(rx − usd)</td>
<td>1</td>
</tr>
<tr>
<td>Israel</td>
<td>dlog(cpi)dlog(gdp)tbr</td>
<td>0.8589</td>
</tr>
</tbody>
</table>

economies. Yet, surveys do not outperform the best VAR on quarterly data. This might be attributable to the fact that there were several structural changes in monetary policy during our sample. On monthly data finance and business surveys perform quite well. The business survey even has a slightly lower forecast error than the best VAR. Household surveys perform worse than the best VAR.\(^{20}\)

For Israel we have only quarterly results up to now. Our sample includes a

---

\(^{20}\)We would like to note that the results on quarterly data use a sample that is one year shorter than the monthly dataset. We are currently redoing the analysis on a longer dataset.
Table 6: Survey errors and errors of the best VAR- monthly data

<table>
<thead>
<tr>
<th></th>
<th>Survey</th>
<th>Best VAR</th>
<th>ARMA</th>
<th>RW</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA spf</td>
<td>1.1825</td>
<td>1.5884</td>
<td>ARMA(3,5)</td>
<td>1.3877</td>
</tr>
<tr>
<td>USA mich</td>
<td>1.5361</td>
<td>1.5432</td>
<td>ARMA(4,4)</td>
<td>1.7172</td>
</tr>
<tr>
<td>Czech Republic- finance</td>
<td>1.9124</td>
<td>1.8985</td>
<td>ARMA(1,3)</td>
<td>4.6753</td>
</tr>
<tr>
<td>Czech Republic-business</td>
<td>1.8951</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech Republic-househ.</td>
<td>2.1569</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Israel</td>
<td>1.5744</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>7.3641</td>
<td>2.0276</td>
<td>ARMA(2,1)</td>
<td>11.0181</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Std. of CPI inflation</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>2.89</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>5.2305</td>
</tr>
<tr>
<td>Poland</td>
<td>13.9382</td>
</tr>
</tbody>
</table>

Table 7: Best VARs– monthly data

<table>
<thead>
<tr>
<th>Variables</th>
<th>α</th>
<th>Lags</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA spf</td>
<td>0.9815</td>
<td>8</td>
</tr>
<tr>
<td>USA mich</td>
<td>0.9827</td>
<td>10</td>
</tr>
<tr>
<td>Czech Republic</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Israel (from 1992)</td>
<td>log(cpi)log(ind)r,b – r_c</td>
<td>1</td>
</tr>
<tr>
<td>Poland</td>
<td>log(cpi) log(ind) u w log(oil) log(m1) r – cb</td>
<td>1</td>
</tr>
</tbody>
</table>

hyperinflationary period. However the survey sample does not include the hyperinflationary episode. Therefore when we are searching for the best VAR forecaster we search it on a sample that does not include the hyperinflationary episode. We find that the best VAR does not use hyperinflationary data. This happens because the treasury bill rate is not available on the whole sample. Therefore we find that if one wants to forecast well after the hyperinflation it is optimal to discard the hyperinflationary data. In Israel surveys do worse then the best VAR.

Finally we would like to note that our surveys are both consumer and professional surveys. Professionals tend to forecast better then consumers. Yet, the Hungarian result shows that even professionals perform worse then econometric forecasts, if the economy unstable. Also in stable economies, consumer surveys perform well compared to econometric predictors. To shed more light on the difference between consumers and professionals we are currently extending our dataset.

We conclude that surveys outperform the best VARs only in countries where the economy is stable. How to define a stable economy is not so clear. We find that the results depend much on how volatile was inflation. Yet regime changes
in monetary policy can also make the environment less stable- or more difficult to forecast- even if inflation has a low variance. If the economy is volatile or there are regime changes surveys do not outperform econometric forecasts.

5 Concluding remarks

In this paper we proposed a test of forecasting performance of surveys: find an econometric predictor that is a good out of sample forecaster, then compare it to survey performance. We have found an interesting pattern. In stable economies surveys can do better then econometric models. In unstable economies surveys provide worse forecasts then econometric models.

We have also shown pitfalls of traditional ways of testing efficiency of expectations. Real time econometric forecasts can fail standard rationality tests. We find that econometric forecasts can be inefficient with respect to variables that are actually used in the econometric model. Conventional efficiency test would conclude that these forecasts were not efficient with respect to these variables. Yet, it might as well be the case that they used these variables efficiently, but the forecast errors are inefficient because there was a change in the data generating process.

If a forecast is inefficient with respect to a macro variable, standard efficiency tests conclude that this variable should have been included in the forecast. We show that this conclusion is misleading. Real time forecasts can become less accurate if one includes this additional variable in the econometric forecast.

We think it is important for future research to examine surveys in different economic environments. This way we can gain a better understanding of how expectations are formed when people face different forecasting environment. The results of this paper suggest that expectations differ in stable and unstable environments. It is a interesting avenue for future research to understand why in unstable economies surveys perform so badly. Is it a result of rational learning in a changing environment? Or is it a result of non rational expectations?
References


Appendix

6.1 Data

Australia
We have quarterly data on CPI from Q1 1950-Q4 2008. AUSTRALIAN $ TO US $ (MTH.AVG.) from Q1 1957. M1 from Q1 1960, M3 from Q2 2004. The central bank interest rate is the cash rate target. Monetary policy decisions are expressed in terms of a target for the cash rate, which is the overnight money market interest rate. This series starts at Q1 1990. Interbank rate (3 month) from Q2 1986. 10 year government bond yield from Q3 1969. Unemployment rate from Q2 1978.

Czech Republic
We use variables from January 1991. CPI, exchange rate (CZECH KORUNY TO US $ ), M2 from January 1993, unemployment rate from February 2001, wage rate


index from January 1992, and industrial production from January 2000. The interest rates we include are: CNB TWO WEEK REPO RATE from December 1995, interbank rate (3 month) from Jan 1994, 10 year government bond yield from April 2000.

Hungary
We use quarterly data from 1q 1995 up to 1q 2006. Other variables are GDP, exchange rate, unemployment rate and treasury bill rate. All variables are available on our sample period.

Israel
We use data from January 1992 to November 2008. We use the CPI data from the Bank of Israel database. Other variables are: M1 (from December 1992), wages (from January 2005), industrial production, exchange rate, ISRAELI SHEKALIM TO $, producer price index of crude oil and natural gas (from January 1005). The interest rates we use are: BANK OF ISRAEL HEADLINE INTEREST RATE (from July 1993), T-BILL YIELD, prime lending rate (from January 1992), and yield on 5 year indexed bonds (from January 1992).

New Zealand
We use quarterly data Q1 1950-Q4 2008. Besides CPI we use the following data: M1 (from Q1 1978), M3 (from Q2 1984), unemployment rate (from Q1 1986), wages (from Q4 1977), industrial production (from Q2 1977), GDP (from Q2 1987), oil price (PPI of input- PETROLEUM, COAL and BASIC CHEMICAL MFG. NADJ from Q2 1994), and exchange rate (NEW ZEALAND $ TO US $ from Q1 1957). We use the following interest rates: central bank target rate (RBNZ OFFICIAL CASH RATE TARGET (OCR) from Q1 1964 ) interbank rate (3 month) (from Q2 1986), prime lending rate (from Q1 1987), and 10 year government bond yield (from Q1 1985).

Poland
Our sample period is January 1991-December 2008. We include the following data: wage, unemployment rate, exchange rate (POLISH ZLOTY TO US $ (AVERAGE)), industrial production, oil price (CPI - ELECTRICITY, GAS and OTHER FUELS from January 1999), M1 (from January 1992), and M3 (from December 1996). The interest rates we used are: central bank discount rate, 3 month money market offer rate, 13 week treasury bill yield (from July 1999), and treasury bill rate (from July 1992).

6.2 Recursive formulation of weighted least squares
In the estimations we use recursive version of weighted least squares. This speeds up the estimation process. In this appendix, we derive the recursive formula cor-
responding to the weighted least squares estimation.

Weighted least squares estimation means minimizing the sum

$$
\sum_{t=1}^{T} \alpha_t (y_t - x_t \beta)^2.
$$

(6)

Here $y_t$ is a scalar, $x_t$ is a row vector, $\beta$ is a column vector. The FOC is

$$
0 = \sum_{t=1}^{T} \alpha_t x_t' (y_t - x_t \beta),
$$

(7)

and therefore the estimate is

$$
\beta_T = \left( \sum_{t=1}^{T} \alpha_t x_t' x_t \right)^{-1} \left( \sum_{t=1}^{T} \alpha_t x_t' y_t \right).
$$

(8)

In this setting it is natural to define

$$
A_T = \sum_{t=1}^{T} \alpha_t, \quad R_T = \frac{\sum_{t=1}^{T} \alpha_t x_t' x_t}{A_T}.
$$

(9)

Thus $R_T$ is a weighted average of the $x_t'x_t$-s. From (9) we have the following recursion for $R_T$:

$$
R_T = \sum_{t=1}^{T-1} \alpha_t x_t' x_t + \alpha_T x_T' x_T = \frac{A_T^{-1} R_{T-1} + \alpha_T x_T' x_T}{A_T}
$$

(10)

From (8) and (9) the recursion for $\beta_T$ is the following:

$$
\beta_T = (A_T R_T)^{-1} \left( \sum_{t=1}^{T} \alpha_t x_t' y_t \right) = (A_T^{-1} R_{T-1} + \alpha_T x_T' y_T)^{-1} \left( \sum_{t=1}^{T-1} \alpha_t x_t' y_t + \alpha_T x_T' y_T \right) =
$$

$$
= (A_T R_T)^{-1} (A_T^{-1} R_{T-1} \beta_{T-1} + \alpha_T x_T' y_T) =
$$

$$
= \beta_{T-1} + (A_T R_T)^{-1} ([A_T^{-1} R_{T-1} - A_T R_T] \beta_{T-1} + \alpha_T x_T' y_T) =
$$

$$
= \beta_{T-1} + (A_T R_T)^{-1} (\alpha_T x_T' y_T - \alpha_T x_T' x_T \beta_{T-1}).
$$

(11)

So the two recursions are

$$
R_T = R_{T-1} + \frac{\alpha_T}{A_T} (x_T' x_T - R_{T-1}),
$$

(12)

$$
\beta_T = \beta_{T-1} + \frac{\alpha_T}{A_T} R_{T-1}^{-1} x_T' (y_T - x_T \beta_{T-1}).
$$

(13)
If we set \( \alpha_t = 1 \), then \( \alpha_T/A_T = 1/T \), and we get least-squares learning. If \( \alpha_t = \kappa^t \) with \( \kappa > 1 \), then
\[
\frac{\alpha_T}{A_T} = \frac{\kappa^T}{1 - \kappa^{T+1}} = \frac{1 - \kappa}{\kappa^{-T} - \kappa}. \tag{14}
\]
If \( T \) is large, this is approximately \((\kappa - 1)/\kappa\), a constant.

If \( \kappa > 1 \) the weight on the first observation \( y_1 \) is \( \kappa \), the weight on \( y_2 \) is \( \kappa^2 \), the weight on \( y_t \) is \( \kappa^t \). The weight on the most recent observation at time \( T \) is \( \kappa^T \).

It is more conventional to have a weight 1 on the last observation. This can be easily obtained if we define \( \alpha = 1/\kappa \), and then multiply each weight with \( \alpha^T \). The the weight on the first observation \( y_1 \) will be \( \alpha^{T-1} \), the weight on the observation \( t \) is \( \alpha^{T-t} \), and the weight on the latest observation \( y_T \) is 1. In this case we have \( \alpha = 1/\kappa = 1 - \gamma \).

In the recursive estimations we calibrate \( \gamma \). For the sake of easier interpretation we report estimation results with the value of \( 0 < \alpha \leq 1 \).