The Additional Information Content of Survey Data

Evidence from Second Moments

MAURIZIO BOVI[•]
ISTAT – Italian National Institute of Statistics
Department of Econometrics and Forecasting
E-mail: mbovi@istat.it

Abstract

Disagreement across forecasters is a well-established fact as widespread is the use of forecasting errors as a proxy of uncertainty. Taking stock of these facts, we aim to test whether the level of disagreement in survey-declared beliefs contains meaningful independent information on the predictive fitness (in terms of squared forecasting errors) of standard econometric models and/or vice versa. Interpreting the estimated squared errors as a proxy of uncertainty we can also shed light on the relationships between the dispersion in survey expectations and macroeconomic volatility. By examining GDP dynamics in the UK we find that the level of discord between expectations revealed in surveys helps improve the predictive power of econometric models and not vice versa. Moreover, the sign of the parameters shows that the greater the entropy in the survey, the higher the uncertainty in the market. On one hand these findings complement previous works, usually focusing on first moments, on the significant one-way information flow running from survey to macroeconomic data. On the other they are in line with - and enlarge our knowledge about - the links between heterogeneity in survey beliefs and aggregate volatility. Results are based on real time data and are robust both to several indicators of dispersion in survey beliefs and to a battery of predictors.

Key Words: Survey Expectations, Forecasting Models, Real-Time Data Analysis.
JEL Classification: C53, D84, E27.

[•] The opinions expressed herein are those of the author and not necessarily reflect the views of ISTAT.

Introduction

The predictive power of survey data not already contained in hard data is often examined and it is easy to understand why. Household survey data i) are available more promptly than national accounts figures; ii) directly elicit people's expectations, a key variable in economics; iii) may register information about the economy that has not yet been reflected in publicly available economic statistics; finally, iv) regardless of how well respondents understand the economy, or how accurate their forecasts have been, if survey data reflect consumers' expectations, they may well help explain agents' behavior. In the literature one can find different points of view about the independence of survey beliefs from economic data. A typical assumption is that agents form model-based beliefs: these latter can not precede realizations. In Carroll's epidemiological approach (Carroll, 2003), for instance, lay individuals update their information set from mass media which, in turn, report economists' forecasts. For the US, Carroll finds that professional expectations on inflation Granger cause those of laypeople and not vice versa. Clearly, on the extent by which lay consumers' expectations are led by forecasts of professionals, one must question the utility of gathering household survey data and the rationale behind the fact that these latter are still among the most watched market movers, even among professional forecasters.[1] Indeed, many empirical papers have been looking, with some success, for the additional information content of survey expectations, where the adjective "additional" means extra-economic factors and/or independent information (for a survey, see Ludvigson, 2004). Keynes' animal spirits or other not-statistically based beliefs, such as the heuristics studied by Kahneman *et al.* (1982), may then impinge on expectations which, in turn, may affect realizations. Katona (1958) has suggested that consumer surveys could capture precisely these kinds of mood-driven expectations. Examining survey data referring to several European countries throughout twenty years, Bovi (2009) shows that agents' expectations exhibit biases significantly conforming to psychological theories which, in turn, are at odds with statistical principles. In sum, there could be some interest in examining the statistical links between survey beliefs and economic data. It should be clear that if the former Granger cause the latter, then survey data would be an important source of extra information about macroeconomic evolutions. It is noteworthy that the literature examining the incremental information content of survey data has typically dealt with first moments.

Due to the undisputed persistence of heterogeneity in beliefs (Evans and Honkapoja, 2001 and 2011; Mankiw *et al.*, 2003; Branch, 2004 and 2007; Lombardelli and Saleheen, 2003; Armantier et al, 2009; Kurz, 2011; Hommes, 2011), some authors have been studying the connections between

[1] Nunes (2009) reports some weaknesses in Carroll's setting.

the second moments of survey data and macroeconomic data (Capistran and Timmermann, 2009; Li and Li, 2010; Badarinza and Buchmann, 2011). It is important to recall that according to a stylized fact dispersion in expectations across forecasters and aggregate volatility are positively correlated. This is not totally surprising given that forecasting fitness measures, such as the mean squared error (MSE), can be thought of as a proxy of volatility.[2] Analyzing the Survey of Professional Forecasters in the USA, Capistran and Timmermann (2009) have shown that macroeconomic uncertainty may lead to disagreement among forecasters. Thus, they argue that information flows run from realizations to beliefs.

Against this framework, our main goal and desired contribution is to test whether lay people use extra information when forming beliefs by examining measures of dispersion. To this end we perform Granger causality and Geweke's instantaneous tests involving MSEs, stemming from standard econometric models, and the level of dissonance in survey responses. This allows to verify whether survey data contain significant information improving the forecasting fitness of econometric models based on hard data only. We examine several models because, to the extent by which people form model-based beliefs, the persistent presence of heterogeneity in expectations implies that they use different predictors. The exercise cannot be, and does not intend to be, exhaustive. First because it is sufficient that in each forecast most, not all, agents use one of the suggested predictors. Then these latter are chosen because i) they are commonly-used; ii) they are contained within the toolbox of a "non-leading" econometrician as, in all probability, the representative citizen populating household surveys would be; iii) the set is an extension of those used in previous works (Williams, 2004; Branch, 2004 and 2007; Forsells and Kenny, 2004; Branch and Evans, 2006); iv) to enrich their representativeness, we estimate the models both recursively and via MSE-minimizing rolling windows. Finally, we compute a forecast combination which, in each period, amounts to take on the most efficient prediction across the whole suite of models under scrutiny. By construction, clearly, this combination generates the best (at least within the models here represented) forecast. In this way we have a really strong benchmark against which to test the extra information content of survey data.

More specifically, once in each quarter we have estimated the models we then use them to perform forecasting exercises. So, we end up with several quarterly time series made up by MSEs (Section 2). We then turn our attention to survey data (Section 3) in order to compute some indicators of the relative disagreement across people's replies (Section 3b). These measures are signal/noise ratios (SNR), based on survey expectations, that are natural second order moments survey counterparts of model-based MSEs. More importantly, a very useful feature of these ratios is that they reduce the

[2] For instance, Kim *et al*., (2004) use the residuals of autoregressive models as a proxy of volatility.

impact of some important issues affecting the basis of widespread methods of quantification of qualitative survey observations. After having studied separately econometric forecasts and survey expectations, we perform Granger causality and Geweke's instantaneous feedback tests to address the significance, the direction and the sign of the time connections between MSEs and SNR (Section 4).

Other two things are noteworthy. First, we take an agnostic view on the direction of the Granger causality linking the second moments of survey data and macroeconomic data, letting the data speak. Second, viewing squared forecasting errors as an indicator of volatility, our analysis can also shed some light on the relationships between the degree of heterogeneity across (non-professional) forecasters' expectations and macroeconomic uncertainty.

Results exhibit a significant and univocal causal chain connecting model-based forecasts and survey-declared beliefs, with the latter preceding the former. This i) supports works claiming that empirical macro models perform better in a variety of dimensions when survey-based expectations are used in place of constructed model-consistent rational expectations (Roberts, 1995, 1997; Mankiw and Reis, 2001, 2003); ii) confirms the additional information content of survey data with respect to hard data by extending evidence to second order moments. Moreover, the sign of the coefficients shows that signal/noise ratios are negatively correlated with the MSEs. That is to say the greater the level of entropy in the surveys, the lower the predictive performance of econometric models. Therefore our evidence also add new knowledge on the above mentioned stylized fact that dispersion in expectations across forecasters and aggregate volatility are positively related. Finally, Geweke's instantaneous feedback tests suggest that there is no significant contemporaneous causality between MSEs and SNR. It turns out that, at least referring to economic growth throughout the period here analyzed, in the UK the expectation feedback system looks like an open loop where survey-declared beliefs play a key role with respect to realizations. These findings are robust to several indicators of dispersion in survey beliefs and to a number of univariate and multivariate predictors estimated both recursively and via rolling windows.

2. The Econometric Framework

2.1 Econometric Models

In this section we describe a battery of commonly used econometric models that we use to predict the GDP annual growth rate in the UK. The logic is that, given the well-established persistence of dispersion in beliefs, if people's expectations are grounded in econometric models then in each (quarterly) forecasting exercise agents should be distributed across several models. The kind of distribution is not important in the present setting: What really matters is that in each quarter most agents use one of the proposed models, which should be likely under model-based beliefs and given the structural dispersion in expectations. The predictor set cannot be, and does not intend to be, exhaustive; each model is chosen for its widespread use in econometric practice and is designed to be representative of a larger class of predictor functions (Branch, 2004). In fact, the models are chosen because they i) are well-known, ii) are contained within the toolbox of a "non-leading" econometrician as, in all probability, the representative citizen populating household surveys would be; iii) constitute a set that is an extension of those used in previous works (Williams, 2004; Branch, 2004 and 2007; Forsells and Kenny, 2004; Branch and Evans, 2006). In addition, the adaptive expectation gain is re-optimized each time (cfr. Table 1) and all AR and VAR models are estimated both recursively and with rolling windows. All that should enhance the representativeness of our selected models.[3] Finally, we compute a forecast combination which, in each period, amounts to take on the most efficient (i.e. MSE-minimizing) prediction across the whole suite of models under scrutiny. It should be clear that this combination generates the best (at least within the proposed models) forecast by definition, resulting in a really strong benchmark against which to test the extra information content of survey data. On the other hand, simpler predictors can be used by less-sophisticated forecasters. Table 1 summarizes the models used in this paper.

[3] Under certain assumptions, the MSE-minimizing window size is equal to the reciprocal of the optimal gain parameter of Kalman's filter estimation (Branch and Evans, 2006). Therefore our results are robust to parameter estimates generated by gain algorithms.

Table 1. The Competing Four-Step-Ahead Predictors

Model (Mnemonic)	*Model Specification (RHS abstracting from error terms)*
Naïve=Random Walk (RW)	y_{t-1}
Adaptive Expectations (AE)	$\gamma_{t-1}y_{t-1} + (1-\gamma_{t-1})AE_{t-1}$ $\gamma_t \in (0,1]$, $\forall t$; starting value = y_{1956Q1}
First Order Autoregression[4] (AR1)	$\alpha_{AR1} + \beta_{AR1}\, y_{t-1}$
Vector Autoregression 1 (VARPC1)	$\alpha_{VARPC1}+\beta_{VARPC1}y_{t-1}+\delta_{VARPC1}\pi_{1,t-1}$ π_1 = Defl. GDP inflation
Vector Autoregression 2 (VARPC2)	$\alpha_{VARPC2}+\beta_{VARPC2}y_{t-1}+\delta_{VARPC2}\pi_{2,t-1}$ π_2 = Defl. consumption inflation
Vector Autoregression 1 (VAR1)	$\alpha_{VAR1}+\beta_{VAR1}y_{t-1}+\delta_{VAR1}\pi_{1,t-1}+\lambda_{VAR1}r_{t-1}$ r = 3M Treasury bill rate
Vector Autoregression 2 (VAR2)	$\alpha_{VAR2} + \beta_{VAR2}y_{t-1} + \delta_{VAR2}\pi_{2,t-1} + \lambda_{VAR2}r_{t-1}$

Note. $y_t \equiv (gdp_t-gdp_{t-4})/gdp_{t-4}$. To mimic the horizon of beliefs elicited in the surveys (Section 3), all models are used to perform four-step-ahead forecasts. Usual information criteria suggest limiting AR and VAR models to one lag. These latter models are estimated each quarter both recursively and with rolling windows. As per γ_{t-1} in the AE formula, it is re-optimized every quarter with step size of 0.04. Due to their symmetry, for VAR models only one of the two, or three, right-hand-side is reported. Details on real time data and estimations can be found in Sections 2.2 and 2.3.

As for the rolling windows estimation variants, we set the smallest window size[5] to thirty-two quarters and the largest to fifty-six quarters. We therefore perform, in each quarter and using the data actually available at the time, twenty-four separate rolling regressions for each model in order to ascertain the optimal (*i.e.* MSE-minimizing) window size.[6] Similarly, a grid search over all $\gamma \in (0,1]$ with step size 0.04 (*i.e.* over twenty-five γ) is performed in every quarter, choosing the value of γ that minimizes the squared error. In each case the period zero initial expectation is the value of y in 1956Q1, which is the very first available data (Section 2.2). This is because all real time data vintages are available from 1956Q1, thus forecasters can choose this date as a starting point in any optimization. This date is sufficiently distant in time from the periods that we study (1985Q1 or later) for any potential initialization issues to be reduced (Carceles-Poveda and Giannitsarou, 2007). It is worth noting that the RW is just one of the twenty-five AE models and, in particular, it is that where *gamma*=1 is arbitrarily imposed in any t (Table 1). Thus, the RW is simpler than, but at its best as good as, the most accurate AE predictor. In this way the AE-family should be well represented in view of our goals.

---

[4] We have also estimated ARMA(1,1) models. Non-reported Wald-tests on coefficient restrictions and usual information criteria show that AR(1) is the best specification for these kinds of models.

[5] In VAR models we must estimate at least four parameters which suggests keeping enough observations (Section 2.2).

[6] We have also estimated more generously lagged (up to four lags) models. Results show that the MSE stemming from AR and VAR models only differing because of their different lag length, are very highly correlated and in no case outperform their most parsimonious version (i.e. that with only one lag).

2.2 Real Time Data

When comparing econometric models and survey expectations, the former should be estimated in real time, i.e., using the data which agents had available at the time when they made their expectations (Croushore, 2011). Taking advantage of the Bank of England database, we can afford to use real time data.[7] Consequently, we do not assume that individuals use information which will be available only in future dates or that they have remarkably good foresight about data revisions (which, in fact, is unlikely). In fact, this caveat also applies to econometric techniques. For instance it is very unlikely that laypeople were able to use VAR models before 1980.

The quarterly data set consists of consecutive vintages ordered as they become available to agents. As for real GDP, the first available vintage was published in March 1976 and covers the period 1955Q1-1975Q4; the second was published in June 1976 and covers the period 1955Q1-1976Q1, and so on up to the last release here used (published in June 2009 and covering the period 1955Q1-2009Q1). Thus, each vintage is released with a delay of three months. In order to perform the standard bi- and tri-variate VAR models discussed in Section 2.1 we need another three variables, namely the deflators of GDP and private consumption,[8] and the interest rate. The first vintages of the two deflators are available since January 1990 and include data from 1970Q1 to 1989Q4. Therefore, time series for deflators are shorter than those for real GDP. This means that univariate real-time forecasting exercises dealing with real GDP dynamics cannot have a starting point before 1976Q4, while the earliest VAR-based expectations must refer to 1990Q4. The interest rate (3M Treasury bill rate) is not subjected to data revision and is available from 1957Q1. To sum up, our database consists of sequentially-released vintages of the above-mentioned four variables.

To mimic the horizon of beliefs elicited in the surveys (Section 3), our real time estimation and forecasting exercises are performed as follows: To obtain the first forecast we estimate the models described in Table 1 using the first available vintage; a quarter later the second vintage becomes available and we re-estimate the above-mentioned models using this new vintage to produce the second forecast, and so on (see also Section 2.3).

2.3 The Forecasting Fitness Measure

To compute the relative predictive accuracy of the models we rely on the MSE statistic. There are several reasons behind this choice. First, the mean squared error is a standard measure of forecasting ability as well as a commonly used loss criterion. Second, it is based on the same statistic that is minimized in least squared estimations, namely the residual sum of squares (RSS):

[7] Data are available on line at: http://www.bankofengland.co.uk/statistics/gdpdatabase/

[8] As far as we know, real time data vintages for CPI are not available.

MSE=RSS/T (T=number of observations). Third, as a measure of dispersion the mean squared error constitutes a natural counterpart of the survey signal-to-noise ratios elaborated in Section 3. Finally, as it is well-known, the Granger causality tests (that we perform in Section 4) are built on the comparison of MSEs.

In all forecasting exercises we compute:

$$MSE_{t+5} = \left({}_{t+5}y_{t+4} - {}_{t+1}y^{e}_{t+4} \right)^2 \qquad (1)$$

where:

$y_t \equiv (GDP_t - GDP_{t-4})/GDP_{t-4}$

${}_{t+1}y^{e}_{t+4}$ = Expected value of y in t+4 based on the vintage released in t+1

${}_{t+5}y_{t+4}$ = Actual value of y in t+4 as reported by the vintage released in t+5.

An example should help to clarify the matter:

in 1976Q1 (=t+1) the first time series for y_t, running from 1956Q1 to 1975Q4 (=t), is made available. The four-steps-ahead prediction refers to 1976Q4 (=t+4). To compare this with its realization we have to wait until 1977Q1 (=t+5), when the actual data for 1976Q4 is eventually released.

As mentioned the sample size is shorter for VAR models, but the logic of the procedure remains the same for all models examined. It is note worthy that the correlations between the MSEs stemming from all models are very high (more than 70%) as are those with actual data. For instance, the correlation between the forecast based on the AE predictor and actual data is as high as 92% and that with the best-combination-model is 96%. It should be suggestive of the fact that should SNR Granger cause best-model grounded MSEs then we can conclude that survey data contain extra information even with respect to a very efficient predictor (based on information grounded in hard data only).

3. Survey Beliefs

3a. *Data*

For our aim, a unique dataset can be obtained from the Business Surveys Unit of the European Commission (European Commission, 1997, 2007). The survey is based on monthly surveys, it starts from January 1985 and it is still running. The survey is not a genuine panel, i.e. there are no re-interviews, though it is designed to and it has accumulated many years of experience in capturing the representative consumer. So, respondents are not professional economists, which is perfectly in

line with our target.[9] Furthermore, even though data are qualitative Pesaran and Weale (2006) have argued that many lay individuals simply provide categorical data even in quantitative surveys. In particular, each nation-wide monthly survey for the UK is based on two-thousand interviews. While the survey asks several questions, the relevant query in the present setting is:
"How do you expect the general economic situation in the country to develop over the next 12 months? It will…".

Surveyed individuals have six reply options:

LB=…get a lot better;
B=…get a little better;
E=…stay the same;
W=…get a little worse;
LW=…get a lot worse;
N=don't know.

LB, B, E, etc., are the shares of respondents having chosen the corresponding option so that they sum up to one. Only these six aggregate shares are available, and only five of them form the basis of this study. Following the mainstream literature, we have excluded the proportion relative to the option[10] "don't know", rescaling the other shares accordingly. Our dataset covers the period January 1985-March 2009. The fact that interviewees must reply to several other questions about both personal (*e.g.*, household financial situation) and macroeconomic evolutions (*e.g.*, unemployment, inflation, *etc.*) may actually induce the respondent to think about GDP when elicited on the general economic situation. In Appendix 1 we report some evidence suggesting that the growth rate closer to what people have in mind when elicited on macroeconomic evolutions is the GDP annual growth. So, in the following sections we limit the analysis to the GDP annual growth rate (y_t). Given that GDP data are available only at quarterly frequency, we aggregate monthly survey data via quarterly averages. The quarterly series obtained using the last or the first monthly observation of the corresponding quarter are almost perfectly correlated to that used in this paper: Our results are robust even to different time aggregations. In view of the analyses performed in Section 4, it is worth mentioning that time aggregation may reduce noise in series, but may also increases contemporaneous correlation (Spencer, 1989).

[9] It is worth noting that, given that we focus on the representative agent in all surveys, demography-related heterogeneous expectations should disappear. In fact, Bryan and Venkatu (2001a,b) and Souleles (2004) detect strong evidence that survey expectations are different across demographic groups.

[10] Possibly, it is a "non response", i.e. it is not the outcome of an explicit elaboration but, rather, a declaration of no information. In this regard, the European Commission Users' Manual (1997, p. 18) states that: "(…) there are six reply options: five "real" ones and a 'do not know' option".

3b. Signal-To-Noise Ratios

In this section we elaborate some alternative quantitative indicators of the mean and dispersion of survey expectations. We do not rely on a single measure because all of them have pros and cons and none has emerged as being definitively superior to the others (Pesaran and Weale, 2006). Needless to say, then, univocal results stemming from different indicators increase the robustness of the findings.

As for the central tendency, one of the most used quantification method is the balance statistic (Anderson, 1952; Theil, 1952). In the three-category scheme (e.g., with only "up", "same", "down" option replies), it is defined as the difference between the share of respondents that expect "up" and the share of respondents that expect "down". Within this approach it can be also computed a measure of disagreement between agents' expectations. To calculate these two moments in the three-category version of the method we, somewhat following Berk (1999), aggregate the EU consumer survey replies. Defining: $s_t^B = LB_t+B_t$; $s_t^W = LW_t+W_t$; $s_t^E = E_t$, the balance and disconformity statistics suggested by Anderson are, respectively:

$$y_t^{e,BAL3} = \alpha(s_t^B - s_t^W) \quad (2)$$

$$\sigma_t^{e,BAL3} = \alpha^2[(s_t^B + s_t^W) - (s_t^B - s_t^W)] \quad (3)$$

We then reckon a slightly modified version, also used by the European Commission (1997), when the survey permits five option replies:

$$y_t^{e,BAL} = \alpha(LB_t+0.5*B_t-0.5*W_t-LW_t) \quad (4)$$

$$\sigma_t^{e,BAL} = \alpha^2[(LB_t+0.5*B_t+0.5*W_t+LW_t)^2 - (LB_t+0.5*B_t-0.5*W_t-LW_t)^2] \quad (5)$$

The parameter α can be chosen to ensure that the balance has the same average value as the GDP growth rate. This is an arbitrary choice and it may be misleading (Nardo, 2003; Pesaran and Weale, 2006). As for the balance statistic, the European Commission (2007) put a unitary weight.[11] Later on we will clarify why we do not address the issue of specifying this multiplier.

Another well-known conversion method is that of Carlson and Parkin (1975, henceforth CP), which we work out in the five option replies version of Batchelor and Orr (1998), too. The CP method interprets the share of respondents as maximum likelihood estimates of areas under the density

[11] In fact, this is the solution usually adopted to publish the index.

function of aggregate expectations, that is as probabilities. The relative mean and standard deviation of expectations across individuals are:

$$y_t^{e,CP} = -\, y_t^r \left(\frac{z_t^3 + z_t^4}{(z_t^1 + z_t^2 - z_t^3 - z_t^4)} \right) \tag{6}$$

$$\sigma_t^{e,CP} = y_t^r \left(\frac{2}{(z_t^1 + z_t^2 - z_t^3 - z_t^4)} \right) \tag{7}$$

where:

y_t^r = agent's reference GDP growth rate;

$z_t^1 = N^{-1}[1-LB_t]$; $z_t^2 = N^{-1}[1-LB_t-B_t]$; $z_t^3 = N^{-1}[1-LB_t-B_t-E_t]$; $z_t^4 = N^{-1}[1-LW_t]$;

$N^{-1}[]$=inverse of the cumulative normal distribution.[12]

With three shares, the first two CP moments are:

$$y_t^{e,CP3} = \delta \left(\frac{N^{-1}[s_t^W] + N^{-1}[1 - s_t^B]}{N^{-1}[s_t^W] - N^{-1}[1 - s_t^B]} \right) \tag{8}$$

$$\sigma_t^{e,CP3} = \left(\frac{2\delta}{N^{-1}[1 - s_t^B] - N^{-1}[s_t^W]} \right) \tag{9}$$

Where δ is a critical value that can be recovered by equating the mean of the expected (or perceived) GDP growth to average actual GDP growth in the sample period. Just like for the multiplier α in the Anderson setting, even the quantification of the critical value is as important as tricky.[13]

Another possible disconformity indicator is the Index of Qualitative Variation (IQV). It is based on the ratio of the total number of differences in the distribution to the maximum number of possible differences within the same distribution:

$$\text{IQV}_t = \frac{K}{K-1} \left(1 - \sum_{i=1}^{K} s_{t,i}^2 \right) \tag{10}$$

[12] Dasgupta and Lahiri (1992), Smith and McAleer (1995), and Berk (1999) find that the accuracy of the quantified series does not significantly vary between any of the common parametric distributions.

[13] For instance, when respondents have five options replies, it is not possible to assume that δ is constant (Pesaran and Weale, 2006).

where K=5 is the number of option replies, and i=LB_t, B_t, E_t, W_t, LW_t. The scaling factor ensures that $0 \leq IQV \leq 1$, where IQV=0 means no variation because all cases belong to a single category, that is to say, expectations are totally homogeneous. Unlike the previous methods, the IQV does not account for the ordered nature of the data. It is important to observe that in our sample (1985:1-2009:1) the mean of IQV was 0.9 with a minimum value as high as 0.73. These figures support once again the fact that heterogeneity in beliefs is as high as long-lasting.[14]

With central tendency and standard deviation indicators at hand, we are eventually able to calculate signal-to-noise ratios aimed to capture the relative heterogeneity in survey expectations. Table 3 shows the definitions of the SNR and their cross-correlations.

Table 3. Survey Signal-to-Noise Ratios and GDP Volatility

		Correlations (1985:1-2009:1)			
Definition	*Mnemonic*	SNR_BAL_t	SNR_BAL3_t	SNR_CP_t	SNR_CP3_t
$y_t^{e,BAL} / \sigma_t^{e,BAL}$	SNR_BAL_t				
$y_t^{e,BAL3} / \sigma_t^{e,BAL3}$	SNR_BAL3_t	0.80			
$y_t^{e,CP} / \sigma_t^{e,CP}$	SNR_CP_t	0.74	0.87		
$y_t^{e,CP3} / \sigma_t^{e,CP3}$	SNR_CP3_t	0.80	0.99	0.89	
$y_t^{e,CP3} / IQV_t$	SNR_IQV_t	0.72	0.95	0.91	0.97

Note: In computing SNR_IQV we have set δ=1. For other definitions see the main text.

Before commenting the figures of Table 3, it is important to note that the proposed SNR afford to overcome some of the difficulties featuring their components. As for the five and three categories probability methods respectively, the factors y_t^r and δ disappear in the ratio. In the balance approach, the remaining α is no more a problem because any linear transformation leaves the correlation unchanged. This said Table 3 emphasizes that, with the possible exemption of SNR_BAL, all measures are strongly correlated. So, notwithstanding the different nature of the indexes, all SNR seem to contain similar information, which is a comforting outcome: it may be the effect of the relatively reduced number of crucial assumptions behind these ratios. Moreover, since our goal of contrasting a survey measure with model-based MSEs, SNR are a natural choice because both criteria deal with second order moments. We also propose a hybrid SNR, namely

[14] Similar results, available upon request, are obtained using the other measures of dispersion here examined.

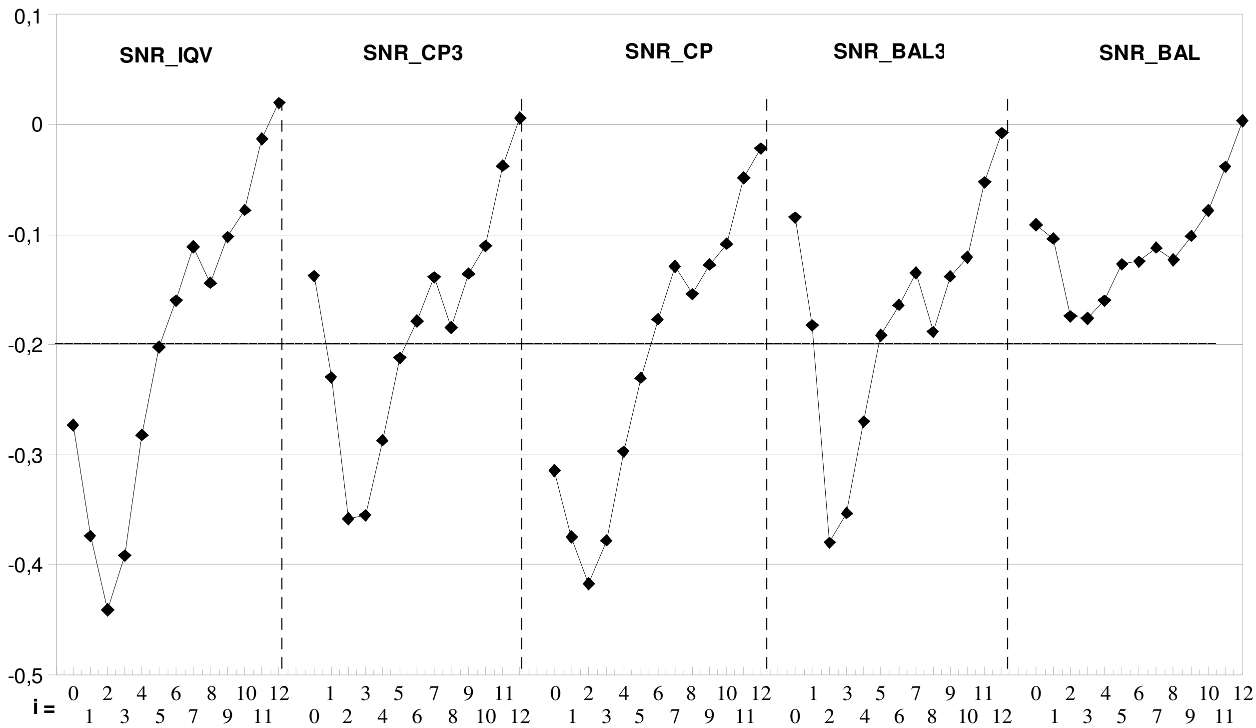
SNR_IQV (computed setting δ=1). The reason is that this latter, according to the Maag's findings (Maag, 2009), might be the best signal-to-noise ratio in order to quantify survey expectations. Relying on micro-data from the Swedish consumer survey that asks for both qualitative and quantitative responses on expected inflation, Maag contrasts the fitness of the CP method with alternative approaches. Evidence suggests that whereas the three-category probability method (CP3) is the best in describing the central tendency, the IQV most closely tracks the actual heterogeneity of quantitative replies. Thus, there are reasons to believe that SNR_IQV could be a good index.

4. Time connections between MSE and SNR

4a. A first look at the data

Prima facie evidence on the links between survey SNR and model-based MSEs is collected in Figure 1. Specifically, it reports the unconditional correlations between contemporaneous MSEs, relative to the AE predictor,[15] and both contemporaneous and lagged SNR.

Figure 1. Model-based MSE and Survey Signal-to-Noise Ratios. Correlations (85Q1-09Q1)



Note: Each point marks the correlation between MSE_AE_t (=MSE deriving from the AE model) and the corresponding SNR_{t-i}, with i=0,…,12. Correlations outside the (approximate) two standard error bounds, equal to ± 0.2 ($=\pm 2/\sqrt{97}$ where 97=n. of obs.), are significantly different from zero.

Figure 1 reveals a number of comforting outcomes and one intriguing result. Among the former, the

[15] The results using MSEs stemming from other models, not reported to save space, are in line with those of Figure 1.

data demonstrates that all significant unconditional correlations are negative. This is comforting because, as mentioned, the MSE can be thought of as a proxy of volatility. As for contemporaneous correlations, the balance measure turns out to be the least accurate. This is to be expected due to the hard-to-defend assumptions behind this method (Batchelor and Orr, 1988). In addition the five-category indicators appear to be superior to the others, perhaps because of their more detailed information content. This is particularly true for SNR_CP and SNR_IQV, which is in line with their above-mentioned relative superiority. Though contemporaneous unconditional correlations are relatively small, they are significant for SNR_CP and SNR_IQV. The intriguing trait of Figure 1 is that *contemporaneous* MSE are significantly correlated with *lagged* SNR.[16] These findings call for more refined statistical analyses that we perform in the next section.

4b. Granger causality

A formal way to examine the correlations reported in the previous section is to estimate bivariate regressions involving one SNR and one MSE:[17]

$$SNR_t = a_0 + a_1 * SNR_{t-1} + \ldots + a_n * SNR_{t-n} + b_1 * MSE_{t-1} + \ldots + b_n * MSE_{t-n} + \varepsilon_t \tag{11}$$

$$MSE_t = a_0 + a_1 * MSE_{t-1} + \ldots + a_n * MSE_{t-n} + b_1 * SNR_{t-1} + \ldots + b_n * SNR_{t-n} + \upsilon_t \tag{12}$$

Where ε and υ are error terms, $n \in [1, \infty)$ and it is selected according to usual information criteria. Working with equations (11) and (12) we perform pairwise Granger causality tests[18] testing $b_1 = \ldots b_n = 0$. Table 5 and 5a summarize the results referring to SNR_IQV.

[16] Unconditional correlations between contemporaneous SNR and lagged MSE, not reported for reasons of space but available upon request, are statistically zero for all lags up to twelve.

[17] Henceforth, to avoid confusion, when we write MSE in capital letter we refer to one of the time series computed in section 2. In fact, the Granger causality is based on the mean-squared error so that when we say that "SNR Granger causes MSE" we mean that the variable SNR helps to reduce the mean-squared forecast error referring to the variable "MSE".

[18] Though Granger block exogeneity tests based on bivariate VARs are another possibility, it was not possible to obtain multivariate normal residuals for all VARs. This said, results from VARs with multivariate normal residuals confirm the outcomes of this paper.

Table 5. Univariate and best model-based MSE vs SNR_IQV ratio. Granger-Causality

*MSE_*	*_RW*		*_AE*		*_AR1*		*_AR1r*		*_MIN*	
Causality Direction	SNR => MSE	MSE => SNR	SNR => MSE	MSE => SNR	SNR => MSE	MSE => SNR	SNR => MSE	MSE => SNR	SNR => MSE	MSE => SNR
Granger Causality	***	ns	***	ns	***	ns	***	ns	**	ns
Sign	-		-		-		-		-	

Note. Granger Causality row reports the levels of significance: "ns" (not significant), * (10%), ** (5%), *** (1%) - based on Newey-West HACSE - of pairwise Granger causality tests. These latter are based on the significance of SNR_IQV terms in the MSE_J equation and vice versa (J=RW, AE, …). The final "r" means that the model is estimated via MSE-minimizing rolling windows. Lag length according to usual information criteria. In order to obtain normal residuals RMSE are used instead of MSE and log(MSE) in the case of MIN(=best-model-combination). The last row reports the sign of the (algebraic sum of the) jointly significant coefficients of SNR entering the MSE equation.

Table 5a. Multivariate model-based MSE vs SNR_IQV ratio. Granger-Causality

*MSE_*	*_VARPC1*		*_VARPC1r*		*_VARPC2*		*_VARPC2r*		*_VAR1*		*_VAR1r*		*_VAR2*		*_VAR2r*	
Causality Direction	SNR => MSE	MSE => SNR	SNR => MSE	MSE => SNR	SNR => MSE	MSE => SNR	SNR => MSE	MSE => SNR	SNR => MSE	MSE => SNR	SNR => MSE	MSE => SNR	SNR => MSE	MSE => SNR	SNR => MSE	MSE => SNR
Granger Causality	***	ns	***	ns	***	ns	***	ns	**	ns	***	ns	**	ns	***	ns
Sign	-		-		-		-		-		-		-		-	

Note. Granger Causality row reports the levels of significance: "ns" (not significant), * (10%), ** (5%), *** (1%) - based on Newey-West HACSE - of pairwise Granger causality tests. These latter are based on the significance of SNR_IQV terms in the MSE_J equation and vice versa (J=RW, AE, …). The final "r" means that the model is estimated via MSE-minimizing rolling windows. Lag length according to usual information criteria. In order to obtain normal residuals RMSE are used instead of MSE and log(MSE) in the case of MIN(=best-model-combination). The last row reports the sign of the (algebraic sum of the) jointly significant coefficients of SNR entering the MSE equation.

Though not reported to save space, the results remain substantially unaffected when using other SNR instead of SNR_IQV. Thus, a robust message emerges from Table 5 and 5a: no matter the survey indicator we use, the S/N ratios always Granger-cause the MSE stemming from several commonly used predictors or their most efficient combination. Otherwise stated, people's relative disagreement on macroeconomic evolutions improves the forecast accuracy of widespread econometric models. Instead, there is no significant information flow in the opposite direction. This causal chain contrasts with the timing behind the epidemiological approach, whereas the "news" coming from econometric models "infects" survey expectations and not vice versa. On the other hand, these findings confirm and complement the evidence on the incremental predictive power of survey data relative to first moments only.

Another strong and intriguing outcome refers to the sign of the correlations connecting SNR to

MSE. The last row of Table 5 and 5a shows that the sign of the coefficients of lagged SNR significantly explaining model-based MSE equation is negative in all the possible combinations between SNR and MSEs. It implies that when the signal coming from the surveys is more perturbed, then the econometric models forecasting ability worsens. This result is in line with the papers of Mankiw and Reis (2001, 2002) and Roberts' (1995, 1997), which show that empirical macro models perform better when survey-based (inflation) expectations are used in place of constructed model-consistent rational expectations. Finally, it is important to note that it has been argued that in Granger causality tests with only one variable measured with error - as, possibly, are our SNR – this latter is often mistakenly concluded as failing to Granger-cause the other variable, while the Granger causality in the other direction is more often detected (Andersson, 2005). That is to say, our results are robust even to measurement errors.

As well-known the definition of Granger causality did not mention anything about the possible instantaneous correlation between time series. It can only establish whether, as found here, past and current values of SNR help to explain the predictive power of econometric models. To analyze the presence of instantaneous correlations in our bivariate regression settings we rely on the Geweke's instantaneous feedback (Geweke, 1982). Simply stated it amounts to test, via log-likelihood ratios, whether the current value of MSE enters the SNR equation (or, equivalently, vice versa): if the added current value significantly enters the equation (11), then there is contemporaneous causality. Table 6 and 6a collect the results.

Table 6. Univariate and best model-based MSE vs SNR_IQV Ratio. Instantaneous Feedback.

Model	*RW*	*AE*	*AR1*	*AR1r*	*MIN*
Geweke's Test	**0.14**	**0.90**	**0.65**	**0.80**	**0.90**

Note. The bottom row reports the χ^2 p-values of LR tests. These latter are the log-ratio made up by the variance of the residuals of the SNR equation of the regression involving SNR_IQV and one the model-based MSE and the variance of the residuals of the same equation where the current value of MSE is added. The null is: "no instantaneous feedback". The final "r" means that the model is estimated via MSE-minimizing rolling windows. See also Table 5 and 5a.

Table 6a. Multivariate model-based MSEs vs SNR_IQV Ratio. Instantaneous Feedback.

Model	*VARPC1*	*VARPC1r*	*VARPC2*	*VARPC2r*	*VAR1*	*VAR1r*	*VAR2*	*VAR2r*
Geweke's Test	**0.20**	**0.72**	**0.25**	**0.51**	**0.04**	**0.50**	**0.02**	**0.28**

Note. The bottom row reports the χ^2 p-values of LR tests. These latter are the log-ratio made up by the variance of the residuals of the SNR equation of the regression involving SNR_IQV and one the model-based MSE and the variance of the residuals of the same equation where the current value of MSE is added. The null is: "no instantaneous feedback". The final "r" means that the model is estimated via MSE-minimizing rolling windows. See also Table 5 and 5a.

The results confirm the relatively low unconditional contemporaneous correlations collected in Figure 1 and almost univocally point to the absence of any contemporaneous feedback between MSEs and SNR.[19] The only two exceptions refer to tri-variate VAR models. Though not crucial for our results, we can speculate that this could be due to the fact that these models are likely to be used by professional forecasters, whose predictions might not to be totally exogenous with respect to contemporaneous survey data. This said, it turns out that not only model-related-MSEs do not Granger cause the degree of heterogeneity pervading survey expectations, we can even conclude that the detected causal chain is stronger than a simple Granger "precedence". All in all, at least referring to economic growth throughout the period here analyzed, in the UK the expectation feedback system looks like an open loop where (survey-declared) beliefs play a key role with respect to realizations.

6. Concluding Remarks

Mainstream economics usually turn a blind eye to the limitations of human rationality. Possibly because of that, we still know little about how people form expectations. Yet, agents' expectations play a pivotal role in economics. In particular, heterogeneous expectations are a fact of life all over the world that seems overlooked by standard economic models.

Using data for the UK, this paper has examined the time connections linking the second order moments of survey-declared beliefs and the forecasting fitness of several standard econometric models. This latter has been measured by squared forecasting errors, which can also be viewed as an indicator of macroeconomic volatility. The relative disagreement in survey expectations has been assessed by unusual signal-to-noise ratios, which allow to tackle some of the issues impinging on commonly used methods of quantification of qualitative survey observations. They are natural survey counterparts of model-based MSEs.

Contrasting with the results obtained for inflation in the US (Carroll, 2003; Capistran and Timmermann, 2009), evidence points to one-way information flows going from survey beliefs to macroeconomic realizations. Specifically, Granger-causality and Geweke's instantaneous feedback tests suggest that model-based MSEs are not a determinant of disparate survey expectations. Instead, divergent expectations significantly affect the forecasting accuracy of macroeconomic models and GDP growth uncertainty. The sign of the correlations points out that higher disagreement in expectations leads to greater macroeconomic uncertainty.

Our findings complement the literature claiming the incremental information content, with respect

[19] As done for Granger causality tests, we have also performed regressions with all the possible combinations stemming from model-based MSEs and the SNR described in Section 3. Results support the evidence reported in this paper.

to that already included in hard data, of central tendencies computed from survey data. We also add evidence on: i) the persistent presence of heterogeneous beliefs, and ii) the positive correlation between dispersion in beliefs across forecasters and macroeconomic uncertainty. Results are robust to several models/ratios combinations, including univariate and multivariate models estimated both recursively and via optimal-size rolling windows.

Appendix 1. Beliefs on next year's general economic situation and GDP growth rates

The correlations collected in Table A1 suggest that the growth rate closer to what people have in mind when elicited on macroeconomic evolutions is the annual growth.

Table A1. Forward-Looking Survey Question and GDP evolutions: Correlations

		Measures of GDP dynamics			
		$a=(gdp_t-gdp_{t-1})/gdp_{t-1}$	$(gdp_t-gdp_{t-4})/gdp_{t-4}$	$[1+a]^4$	Output gap
Survey Central Tendency	$y_{t-4}^{e,BAL}$	0.14	0.26	0.14	0.22
	$y_{t-4}^{e,BAL3}$	0.11	0.21	0.11	0.20
	$y_{t-4}^{e,CP}$	0.14	0.29	0.14	0.22
	$y_{t-4}^{e,CP3}$	0.16	0.28	0.16	0.27

Note: Central tendency indicators, all computed with unitary multipliers, are defined in Section 3b. They are lagged because the survey question refers to one-year-ahead GDP growth: "How do you expect the general economic situation in the country to develop over the next 12 months?"
Output gap=Log(gdp/gdp*); gdp*=Hodrick-Prescott GDP trend (λ=1600). All dynamics deal with real time vintages.

Given the wording of the questions faced by the interviewed (Section 3), evidence suggesting that survey data are more related to the annual growth rate than to other GDP dynamics is expected and, hence, somewhat reassuring about the reliability of the answers. In fact, this is the rate usually chosen by other works in this topic (e.g. Branch, 2004 and 2007; Capistran and Timmermann, 2009). By the same token, another expected and reassuring outcome is that correlations referring to backward-looking queries (Table A2) are larger than those referring to forward-looking questions (Table A1). More importantly, the GDP annual growth again shows the higher correlation with SNR.

Table A2. Backward-Looking Survey Question and GDP evolutions: Correlations

		Measures of GDP dynamics			
		$a=(gdp_t-gdp_{t-1})/gdp_{t-1}$	$(gdp_t-gdp_{t-4})/gdp_{t-4}$	$[1+a]^4$	Output gap
Survey Central Tendency	$y_{t}^{e,BAL}$	0.47	0.64	0.47	0.55
	$y_{t}^{e,BAL3}$	0.44	0.61	0.44	0.54
	$y_{t}^{e,CP}$	0.53	0.69	0.52	0.58
	$y_{t}^{e,CP3}$	0.54	0.70	0.54	0.62

Note: Survey question: "How do you think the general economic situation in the country has changed over the past 12 months?". See also under Table A2.

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.

Andersson, J., 2005. Testing for Granger causality in the presence of measurement errors, Economics Bulletin 47, 1−13.

Armantier, O., Bruine de Bruin, W. Topa, G. van der Klaauw, W. and Zafar, B. Inflation Expectations and Behavior: Do Survey Respondents Act on Their Beliefs? *Staff Report*, Federal Reserve Bank of New York, 509.

Badarinza, C. and Buchmann, M., (2011). Macroeconomic vulnerability and disagreement in expectations in the euro area, *ECB Working Paper* N. 1407.

Batchelor, R.A. and A.B., Orr. (1988): Inflation Expectations Revisited. *Economica*, 55 (219), 317-331

Berk, J. M. (1999): "Measuring Inflation Expectations: A Survey Data Approach," *Applied Economics*, 31(11), 1467-1480.

Branch, W.A., (2004): "The theory of rationally heterogeneous expectations: evidence from survey data on inflation expectations". *Economic Journal*, 114, 497.

Branch, W.A., (2007): "Sticky information and model uncertainty in survey data on inflation expectations". *Journal of Economic Dynamics and Control*, 31, 245–276.

Branch, W.A., Evans, G.W., (2006): "A simple recursive forecasting model". *Economics Letters,* 91, 158-166.

Bryan, M.F., Venkatu, G., (2001a): "The demographics of inflation opinion surveys". *Economic Commentary*, Federal Reserve Bank of Cleveland.

Bryan, M.F., Venkatu, G., (2001b): The curiously different inflation perspectives of men and women. *Economic Commentary,* Federal Reserve Bank of Cleveland.

Carroll, C.D., (2003): "The epidemiology of macroeconomic expectations". *Quarterly Journal of Economics*, 118, 1.

Capistran C and A. Timmermann (2009). Disagreement and biases in inflation expectations. *Journal of Money, Credit, and Banking*, 41:365–396.

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

Croushore, D., 2011. Frontiers of Real-Time Data, Journal of Economic Literature 49, 72-100

Dasgupta, S. and K. Lahiri, (1992): "A Comparative Study of Alternative Methods of Quantifying Qualitative Survey Responses Using NAPM Data, *Journal of Business and Economics Statistics*, 10, 391–400.

Geweke J. (1982): “Measurement of linear dependence and feedback between multiple time series”, Journal of American Statistical Association. 77, 304–313.

European Commission (2007) *The Joint Harmonised EU Programme of Business and Consumer Surveys, User Guide,* European Commission, Directorate-General for Economic and Financial Affairs, July.

Evans, G.W., Honkapohja, S., 2001. *Learning and Expectations in Macroeconomics.* Princeton University Press, Princeton, NJ.

Evans, G.W. and Honkapohja, S., 2011. Learning as a Rational Foundation for Macroeconomics and Finance, Bank of Finland Research Discussion Paper, N. 8/2011

Forsells, M., and G. Kenny (2004): Survey Expectations, Rationality and the Dynamics of Euro Area Inflation," *Journal of Business Cycle Measurement and Analysis*, 1(1), 13-41.

Kim, C.J., C.R. Nelson and J. Piger (2004): “The Less-Volatile U.S. Economy: A Bayesian Investigation of Timing, Breadth, and Potential Explanations”, *Journal of Economic and Business* Statistics, 22, 80-93.

Li, D. and G. Li, (2010). Belief dispersion, dispersion of belief changes, and trading volume: Evidence from surveys of consumers and professional forecasters. Working paper available at SSRN: http://ssrn.com/abstract=1572439.

Ludvigson, S. (2004): “Consumer Confidence and Consumer Spending”, *Journal of Economic Perspectives,* 18**,** 29-50.

Maag, T. (2009): “On the Accuracy of the Probability Method for Quantifying Beliefs about Inflation,” *Working papers* 09-230, KOF Swiss Economic Institute, ETH Zurich

Mankiw, N. G., and R. Reis. (2001): “Sticky Information: A Model of Monetary Nonneutrality and Structural Slumps,” *NBER Working Paper Number 8614*.

Mankiw, N. G., and R. Reis, (2003): “Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve.” *Quarterly Journal of Economics* 117 (4), 1295-1328.

Mankiw, N. G., Reis, R.,Wolfers, J., (2003): “Disagreement about inflation expectations”, *NBER Macroeconomics Annual.*

Nardo, M. (2003): “The quantification of qualitative survey data: A critical assessment”. *Journal of Economic Surveys*, 17:645–668.

Newey W.K. and K.D. West (1987): “A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix”, *Econometrica*, 55, 703-708.

Pesaran M.H. and M. Weale (2006): Survey expectations. In G. Elliot, C. W. J. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*, volume 1, chapter 14, pages 715–776. Elsevier.

Roberts, John M. (1995): “New Keynesian Economics and the Phillips Curve,” *Journal of Money, Credit, and Banking*, 27(4), 975–984.

Roberts, John M. (1997): “Is Inflation Sticky?,” *Journal of Monetary Economics*, pp. 173–196.

Souleles N.S. (2004): “Expectations, Heterogeneous Forecast Errors, And Consumption: Micro Evidence Form The Michigan Consumer Sentiment Surveys,” *Journal of Money, Credit and Banking*, 36, 39-72.

Spencer, David E. (1989): “Does Money Matter? The Robustness of Evidence from Vector Autoregressions,” *Journal of Money, Credit and Banking*, 21, 4, 442-54.

Smith, J., and M. McAleer (1995): Alternative Procedures for Converting Qualitative Response Data to Quantitative Expectations: An Application to Australian Manufacturing," *Journal of Applied Econometrics*, 10(2), 165-185.

Theil, H. (1952): “On the Shape of Micro-variables and the Munich Business Test,” *Revue de l'Institut International de Statistique*, 20, 105-120.