The utility of expectational data: Firm-level evidence using matched qualitative-quantitative UK surveys*

Silvia Lui, James Mitchell and Martin Weale

National Institute of Economic and Social Research

November 10, 2009

*Address for correspondence: s.lui@niesr.ac.uk; j.mitchell@niesr.ac.uk and m.weale@niesr.ac.uk. We thank the CBI and the ONS for their help in facilitating this project, with particular thanks to Lai Co (CBI), Rhys Davies (ONS), Robert Gilhooly (ONS), Felix Ritchie (ONS), Eric Scheffel (ONS), Richard Welpton (ONS) and Jonathan Wood (CBI). We gratefully acknowledge financial support from the ESRC (Award Reference: RES-062-23-0239) and the Bank of England. For more details about the CBI's ITS, please go to http://www.cbi.org.uk/ndbs/content.nsf/802737AED3E3420580256706005390AE/2F172E85D0508CEA80256E20003E95C6.

Abstract

This paper assesses the utility of qualitative expectational survey data at the firm-level in terms of both their ability to anticipate firms' subsequent retrospective, but qualitative, reports of their performance but also these same firms' quantitative answers. The assessment requires access to a unique panel dataset which matches firms' responses to a leading qualitative tendency survey conducted by the Confederation of British Industry with these same firms' quantitative replies to a different survey carried out by the Office for National Statistics. We employ nonparametric tests of the so-called "best-case scenario" and introduce a weaker test for the coherence between these two surveys and test whether the qualitative data contain a(ny) signal about the quantitative data. We find that while firms' qualitative expectations are "best-case" predictions of their qualitative assessment of their output growth they do not contain a signal about the quantitative data. But we can reject the null hypothesis of noise for the retrospective qualitative data. We discuss this apparent paradox and suggest that qualitative business survey data are more useful for nowcasting than forecasting.

1 Introduction

Expectational data continue to be used widely as leading indicators by forecasters. But doubts persist about the quality and utility of these expectational data. This is explained, at least in part, by the qualitative nature of data typically available from business tendency surveys. Tendency surveys ask firms (or consumers) about the expected direction of change for some economic variable: the variable is expected to 'go up', 'stay the same' or 'go down'. This has impeded comparison of the expectational data against the subsequent (quantitative) realisations data and thereby hindered the establishment of an empirical consensus on the utility of expectational data.

The majority of applied studies assessing the informational content of expectational data have first quantified and aggregated the underlying firm or consumer-level qualitative survey data and then related this survey-based aggregate to economy-wide output growth or inflation, say (e.g., see Driver & Urga (2004)). Moreover, this quantification and aggregation is usually implicit, as the expectational data are usually made available only in processed form as the proportions of optimists and pessimists or the difference between these two proportions - the so-called balance of opinion. (See Pesaran & Weale (2006) for a survey of the various quantification approaches.) For example, the cross-country expectational business and consumer survey data archived by DG-ECFIN at the European Commission take this form.

Nevertheless a small but growing literature has peered into the 'black box' to examine the underlying firm-level qualitative panel data. When these data take the form of a panel, such that it is possible to keep track of the expectations and realisations as reported by individual firms, then direct tests of the quality of the qualitative survey data can be employed. Nerlove (1983) compared French and German firms' qualitative expectations with their subsequent qualitative reports of what actually happened using 3×3 contingency tables constructed from their trichotomous ordered responses; and Horvath et al. (1992) and Ivaldi (1992) tested rationality using the (polychoric) correlation matrix between the categorical variables. Mitchell et al. (2005) considered how to construct indicators of the aggregate of interest from the qualitative panel data which gives more emphasis to firms whose answers have a close link to the official data than to those whose expectations correspond only weakly or not al all.

This paper seeks to extend the coverage of empirical work by assessing the usefulness of these qualitative expectational data at the firm-level not just in terms of their relationship with firms' own retrospective, but qualitative, reports of their output growth but to these same firms' quantitative answers. Some surveys, like the annual Dutch Socio-Economic Panel analysed by Das et al. (1999), in addition to qualitative expectational data provide both qualitative and quantitative retrospective assessments of, in this case, households' incomes. But it is far more common for the higher (monthly) frequency business surveys of the sort used by forecasters (such as the DG-ECFIN data), because of their timeliness, to provide only qualitative data. This has hitherto prevented assessment at a firm-level of the qualitative expectational data against the quantitative realisations data.

Therefore for this study we construct a matched dataset which relates UK manufacturing firms' qualitative responses to the Confederation of British Industry's Industrial Trends Survey to these same firms' quantitative assessments of their own performance which they supply to the Office for National Statistics (ONS) as part of its Monthly Production Inquiry (MPI). The MPI serves as one basis for the construction of the national accounts. The CBI's survey is the basis for the UK survey held in the European Commission's archive of business surveys - a common source of information for forecasters (and used in many studies, such as Claveria et al. (2007)). The two surveys are based on different sampling and data measurement methodologies. But given that both surveys elicit responses about variables with similar names, economists routinely use the more timely qualitative surveys to infer the quantitative survey; e.g., see Ashley et al. (2005).

We then assess the utility of both the retrospective and prospective qualitative survey questions by comparing them at the firm-level against the quantitative realisations data. We follow Das et al. (1999) and first assess the utility of the qualitative expectational data using non-parametric tests of the "best-case scenario" that firms have rational expectations and that reported expectations are *optimal* predictions of future outcomes, in the sense that they minimise the firm's loss function. Under the best-case scenario, restrictions or bounds on the distribution of realised outcomes can be derived conditional on firms' qualitative expectations. We will see that while qualitative expectational data, even under the best-case scenario, cannot tell us exactly what is going to happen they do imply bounds. Manski (1990) previously studied this problem for binary, rather than ordered, survey responses. Importantly, as it is unclear how firms reply to the qualitative surveys, tests are carried out assuming the firms have either the mean, mode or median (more generally the α-quantile) of their subjective distribution in mind when reporting their *optimal* categorical and ordered expectation.

We also test the *coherence* between the retrospective qualitative data and the quantitative realisations data. This is important, given that these data are from different surveys, with different sampling and data measurement assumptions. The widespread use of qualitative business surveys, like the CBI's, to provide both more timely estimates of the quantitative data (using the retrospective responses) and forecasts (using the prospective, expectational data) is implicitly predicated on the assumption that the two samples are measuring the same concept of output growth.

The plan of the remainder of this paper is as follows. Section 2 introduces the timely qualitative business survey and the less timely but quantitative survey by the ONS and discusses how firms' responses across the two surveys were matched. Section 3 then considers and reports the nonparametric tests of the best-case scenario when firms reply to the qualitative survey with the mode, median or mean of their subjective density forecast in mind. For the mean, we also introduce and implement a test of the coherence between the qualitative and quantitative data. This tests whether the qualitative survey data contain a signal about the quantitative data or whether they are simply noise. Section 4 concludes.

2 The matched qualitative-quantitative panel

2.1 Background

The two underlying surveys of interest are the ONS's Monthly Production Inquiry (MPI) and the CBI's Industrial Trends Survey (ITS). The MPI is a sample-based survey covering all of the UK and is the basis for the monthly Index of Production (IoP), a key 'national statistic' in the UK. In the MPI the ONS asks close to 9000 firms each month for quantitative information on their turnover values in the month. (Appendix 1 provides additional details on the two surveys and their sampling methodologies). The aggregated (across firms) responses of the MPI are then used to construct the monthly IoP. The firm-level MPI turnover data are deflated, using the IoP deflator at the 4 digit level, to produce figures in volume terms. Output change, y_{it}, is then defined as firm i's turnover in month t relative to its turnover in month $(t-3)$.[1] In addition, 5% Winsorisation on each of the upper and lower tails of the distribution of the output growth rates, y_{it}, pooled across firms for each period, is carried out to mitigate the possible effects of outliers.

[1]The 3-month reference period is motivated, as we discuss below, by the reference period over which firms are asked to report when replying to the CBI's qualitative survey and the need to ensure hard data on the reference (outturn) series are not known (and in the forecaster's information set) for one or more of the 3-months to which the reference series relates. Results were robust to different assumptions about the length of this window.

Two-tailed Winsorisation (Dixon 1960) involves replacing those values of a variable below the lower or above the higher x-percentile with the values observed at those percentiles. It is generally preferred to trimming as a means of dealing with outliers. We discuss the effects of Winsorisation as we present our results.

The MPI questionnaires are sent out to firms three days before the end of a calendar month; the majority of firms (the MPI achieves over a 80% response rate at the time of publication) then reply 18 working days into the following month. The IoP, the aggregation of these firm-level data, is then published on the 26th working day after the end of the reference month with the consequence that it is published more than a calendar month after the month to which it relates; see ONS (2005) for details.

The ITS is a voluntary survey open to both CBI members and non-member companies. It asks about 900 firms each month many questions, only some of which are 'verifiable' (i.e., testable against official data). In this paper we focus (i) on the retrospective question, "Excluding seasonal variations, what has been the trend over the past four months with regard to volume of output?"; and (ii) the prospective question, "Excluding seasonal variations, what are the expected trends for the next four months with regard to volume of output?". Firms reply "up", "same" or "down" (and "not applicable").[2] The ITS is published at the end of the month concerned, and it therefore gives an impression of being more than a calendar month ahead of the MPI. This explains the potential value of the ITS, and in particular the first question, as a timely source of (nowcasting) information about the current state of the economy. Moreover the second question, which is forward-looking, has no counterpart in official surveys. However, firms fill in their ITS forms between the beginning of the last week of the preceding month and the middle of the current month. As a result the ITS does not cover all of the month in which it is published. Coupled with a longer reference period (as firms in the ITS are asked about the last three/four months) it becomes apparent that while the ITS is published ahead of the MPI it need not contain more timely information about economic activity in the current month.

The MPI asks firms about their turnover while the ITS asks about output. The difference between these is accounted for by the change in stocks of finished goods and work in progress. This means that the comparison that we make is between monthly sales as reported to the ONS and the response to the ITS which should indicate what is happening to output. When fluctuations in sales growth are not met from stocks, and

[2]Since June 2003, firms were, in fact, asked for their trend over the last three months. This change does not affect our results.

lead directly to output movements, the MPI and the ITS responses should have a close direct relationship.

No information is available on precisely who, at a given firm, fills out the survey form nor on how often this person changes over time. The CBI survey is generally replied to by a board member, while the ONS survey, at least for larger firms, may be filled in at a lower level.

We relate the prospective qualitative data at month $(t-4)$ with the retrospective qualitative data at month t. But, as indicated above, we define the quantitative retrospective data, y_{it}, over the interval t to $(t-3)$ to ensure that when firms' report their expectations (for the next three/four months) at month $(t-4)$ they definitely do not know any of the quantitative realisations data (the outturn they are trying to forecast) when they form their expectations. In any case, as mentioned above, results were robust to playing with these timing assumptions.

2.2 The matched dataset

To match firms' responses across the two surveys we arranged for the CBI to provide the data collected by the ITS to the ONS in a manner which preserved obligations of confidentiality for both bodies. The CBI explained to its respondents that they would remain anonymous to users of the data and gave them the chance to opt out of having their responses matched to the ONS data. The data set available to us began in 2000 because the CBI changed its coding system at the start of 2000 and it was not possible to link data collected in 1999 and earlier with those collected from 2000 onwards. The CBI advised that respondents' co-operation was more likely if the data passed to ONS were not very recent and on these grounds asked its respondents to agree to the provision of data for the period 2000-2004. Only five respondents out of 2589 declined.

The ONS then matched the CBI data set to its own MPI for the five years 2000-2004, inclusive. This was achieved by text matching based on common variables, specifically the names, addresses and postcodes of firms. Of the 2584 different firms who replied at least once to the ITS, 807 different firms gave at least one contemporaneous matched response to the ITS and MPI over the five years. Over the five years, as 807 different firms make the matched ITS-MPI dataset and 2584 different firms were sampled at least once by the CBI, the match rate against the ITS is 31%. On average 170 different firms are present in the matched dataset each month. The total number of times a given firm, firm i, is present in the matched dataset, T_i, ranges from 1 to 60. The average number (across firms) of matched time-series observations is 12.7, i.e. $\overline{T_i} = 12.7$. 25% of firms have at most 3 matched (contemporaneous) responses; 50% of firms have at most 8 matched responses and 75% of firms have at most 16 matched responses. We refer the reader to Lui et al. (2008) for more details of the statistical properties and representative-ness of the matched dataset. Here we simply note that the matched dataset over-represents large firms. This is explained, in large part, by the stratified nature of the MPI sample which ensures that large firms are more likely to remain in the sample. But there is no reason to suspect that the matched dataset picks up firms that are either particularly *good* or *bad* at forming expectations. What matters is whether the relationship between the ITS and MPI data, at a firm-level, depends on the probability that a firm is in the matched dataset; in other words, whether the relationship depends on the type of firm under consideration. In any case, the bias towards large firms is not in itself a problem since a finding that the ITS and the MPI responses were not related in our matched sample would be of considerable interest even if it applied only to large firms and nothing were known about small firms.

The nature of these matched data and in particular the fact that the panel is almost certainly too incomplete to be described as a (balanced) panel, motivates our focus on nonparametric means of testing the utility of the expectational data. Access to the matched dataset, as we discuss below, is required when we wish to evaluate the utility of the expectational data relative to the quantitative realisations data. When, as is more common, evaluating relative to firms' qualitative assessments of their realisations we can use the ITS alone. While the ITS has an average sample size each month of about 900 firms, about 540 firms provide responses to both the time t and time $(t-4)$ surveys enabling us to match up their qualitative expectations at time $(t-4)$ with their subsequent realisations reported at time t.

3 Nonparametric testing of qualitative survey data

The retrospective and prospective qualitative survey data provide two pieces of categorical information on the firm-specific random variable y_{it}, which denotes firm i's output growth rate in time t (relative to $t-3$):

1. a prediction of y_{it} made at the end of period $(t-4)$. The prediction is denoted by the discrete random variable $y^p_{it,j}$, $j = 1, 2, 3$ (corresponding to "down", "no change"

and “up”, respectively), where

$$y^p_{it,j} = 1 \text{ if } a_{j-1,it} < y^*_{it} \leq a_{j,it}; 0 \text{ otherwise} \tag{1}$$

2. the actual outcome in period t. The outcome is denoted by the discrete random variable $y^r_{it,j}$, $j = 1, 2, 3$ where

$$y^r_{it,j} = 1 \text{ if } b_{j-1,it} < y_{it} \leq b_{j,it} \text{ ; } 0 \text{ otherwise} \tag{2}$$

We follow convention and assume $\{a_{0,it}, b_{0,it}\} = -\infty$ and $\{a_{3,it}, b_{3,it}\} = \infty$. Note that the thresholds can vary across firms, i and across time, t.

Official quantitative data on y_{it} are available from the MPI and are published 26 working days after the end of month t.

Let the following 3×3 contingency table summarise the probability distribution characterising $y^p_{it,k}$ and $y^r_{it,j}$:

$$\begin{array}{cc|ccc} & & & y^p_{it,k} & \\ & j/k & 1 & 2 & 3 \\ \hline & 1 & p_{11,t} & p_{12,t} & p_{13,t} \\ y^r_{it,j} & 2 & p_{21,t} & p_{22,t} & p_{23,t} \\ & 3 & p_{31,t} & p_{32,t} & p_{33,t} \end{array} \tag{3}$$

where $p_{jk,t}$ denotes the probability at time t of observing realisation j and expectation k; $\sum_{j,k} p_{jk,t} = 1$, $p_{jk,t} > 0, j, k = 1, 2, 3$. If at time t a sample of N_t observations is drawn from a population characterised by the above distribution then let $n_{jk,t}$ denote the counts in category j, k, where $\sum_{j,k} n_{jk,t} = N_t$, and $(n_{jk,t}/N_t) \xrightarrow{p} p_{jk,t}$. The analysis below ignores effects arising from the sample design of both the MPI and ITS.

3.1 Rationality: the “best-case scenario”

Tests of the rationality of expectations remain a useful first step in establishing the utility of expectational data. As rationality is commonly rejected at the macroeconomic (aggregate) level we might expect it to be rejected at the firm-level also, since rationality on average (across firms) is a weaker condition than firm-level rationality (e.g., see Pesaran & Weale (2006)). But even when irrational (in the strict sense), as we discuss below, the expectational data may still be “useful”, in the sense that they may still contain a signal about the quantitative realisations data.

If rational, the firm's subjective density forecast equals the 'objective' density function of the realisation. The realisation y_{it} is therefore drawn from the same density as the forecast. To test rationality we compare the expectational data with the distribution (across firms) of both quantitative and qualitative realisations data. The rationality tests proposed by Das et al. (1999) rely on the assumption that the realisations data are independent across firms. (Note that the expectational data are independent given random sampling.) Thereby, we exclude the possibility of macroeconomic (common) shocks, which would induce dependence in the realisations data across firms. The fact that the rationality tests are conditional on the auxiliary assumption of independence, motivates Das et al. (1999) to consider the tests as tests of the "best-case scenario". If the best-case scenario is unrealistic then the expectational data contain (even) less information than the bounds imply. Below we consider the independence issue further, and for quantitative realisations data consider how ex post one can identify common (macroeconomic) shocks and control for them when testing the rationality and utility of the expectational data.[3]

To use the qualitative realisations data $y^r_{it,j}$ Das et al. (1999) need to assume $a_{j,it} = b_{j,it}$. Without this assumption, in the absence of quantitative information on these thresholds, the best-case scenario cannot be tested for $y^r_{it,j}$ and it is impossible to compare the qualitative realisations and expectational data.

The best-case scenario tests considered below also depend on which feature of the firm's subjective density forecast is reflected by $y^p_{it,k}$. This depends on a firm's loss function. Following Das et al. (1999) we distinguish between the cases when $y^p_{it,k}$ depends on the mode, median (or, more generally, the α-quantile) or the mean of the firm's density forecast. While tests of the best-case scenario based on the mode and median are constructed based on the (retrospective and prospective) qualitative survey data only, to test under the mean requires knowledge of the quantitative realisations data and the thresholds. We also introduce a weaker test, that does not require knowledge of the thresholds, of whether the expectational data contain a signal or are simply noise. We then test whether the qualitative data are coherent with the quantitative data. This is

[3]Tests which make less restrictive assumptions are possible when quantitative expectational data, perhaps even available as density forecasts, are available also. However, as indicated in the introduction, the majority of surveys continue to elicit qualitative expectations only; since the results of these surveys are used widely it remains of interest to assess the quality of the data, at the underlying firm-level, even though the tests which have to be employed inevitably have to make additional assumptions in order to be operational.

important given that results from qualitative surveys are used routinely as a means of adding strength to quantitative surveys since they are typically published at a greater lag. These tests therefore distinguish two cases for the realisation, y_{it}; in the first it is observed quantitatively, and in the second we observe only the category $y^r_{it,j}$ $(j = 1, 2, 3)$ in which y_{it} is contained.

3.1.1 Modal category assumption

Under rationality, and when firms base their categorical expectation $y^p_{it,k}$ on the mode of their subjective density forecast,

$$p_{kk,t} \geq p_{jk,t};\ j = 1, 2, 3 \tag{4}$$

The best-case scenario implies that for any group of firms who report $y^p_{it} = k$ more realisations subsequently fall in category k than the other categories. The same condition for rationality is, in fact, also derived by Gourieroux & Pradel (1986); although they assume that rather than reporting the mode, firms set their expectation to minimise squared error loss defined in terms of the 1-norm distance ($y^p_{it,j}$ and $y^r_{it,j}$ may only take on the values 0 and 1) between $y^p_{it,j}$ and $y^r_{it,j}$.

This condition, (4), which involves comparison of the probabilities in a given column of the contingency table (3), can be compared with self-fulfilling expectations, which involve comparison across rows, and imply that what is expected subsequently happens so that $p_{kk,t} > p_{kj,t}$.

A test of the null hypothesis that $p_{kk,t} \geq p_{jk,t}$ versus the one-sided alternative $p_{kk,t} < p_{jk,t}$ can be constructed using

$$\sqrt{n_{.k,t}} \left(\frac{(\widehat{p}_{kk,t} - \widehat{p}_{jk,t})}{\sqrt{2\widehat{p}_{kk,t}}} \right) \rightarrow N(0, 1) \tag{5}$$

where $n_{.k,t} = \sum_{j=1}^{3} n_{jk,t}$ and $\widehat{p}_{jk,t}$ is a sample estimate of $p_{jk,t}$, given that, under independence of $y^r_{it,j}$ across i, $\widehat{p}_{kk,t}(1 - \widehat{p}_{kk,t})$ is the variance of $\widehat{p}_{kk,t}$ and the covariances are $-\widehat{p}_{jk,t}\widehat{p}_{j'k,t}$ $(j \neq j')$.

Figure 1 plots the number of firms present, across time, in the ITS. It shows that on average 540 firms are present each month, although there is some variation across time in the number of respondents. Figure 2 then plots, across time, the proportion of these firms that replied up (U), same (S) or down (D) given that they expected an up

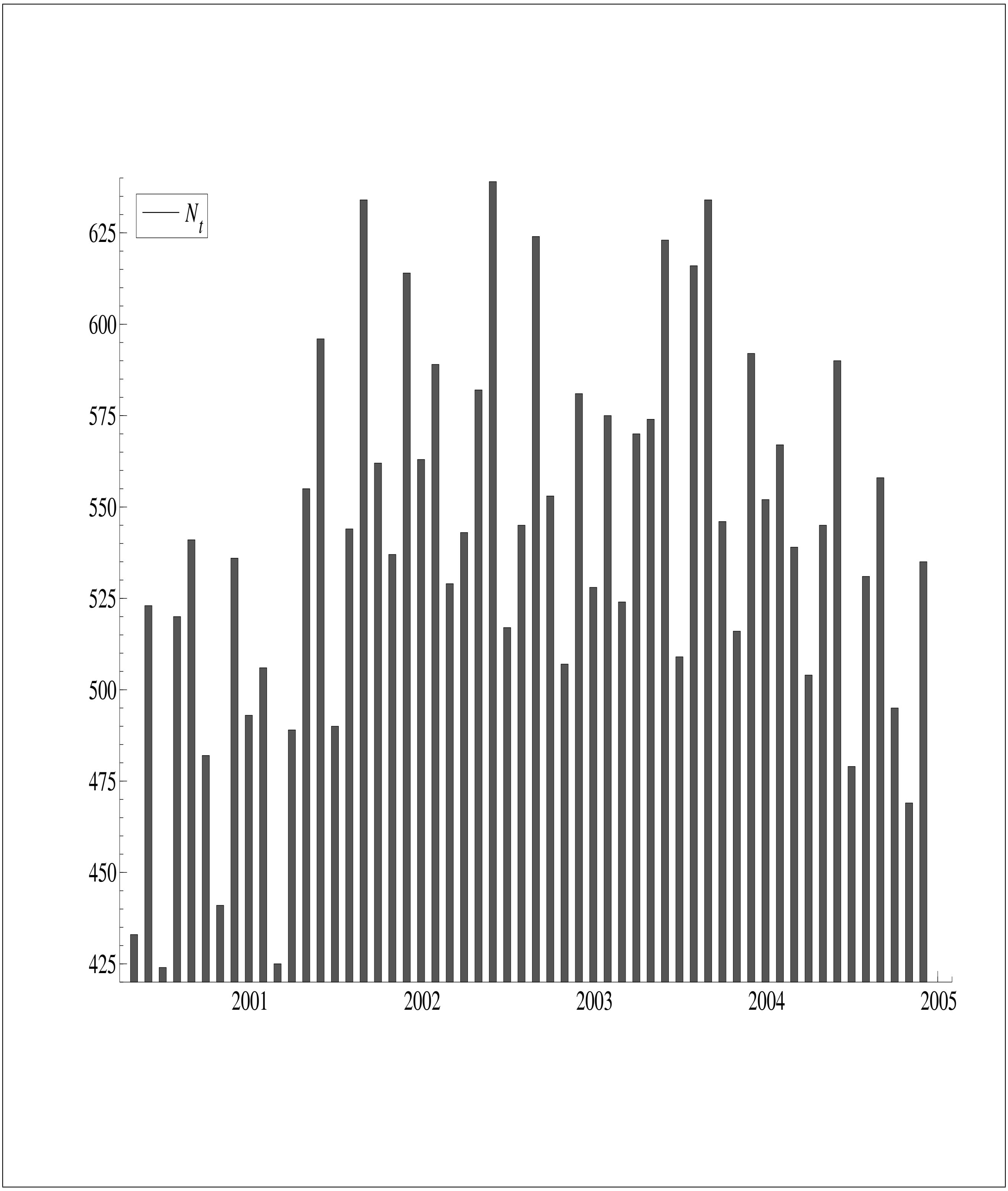

Figure 1: Number of firms in the ITS at month t that give both a retrospective response at time t and a prospective response at time $(t-4)$

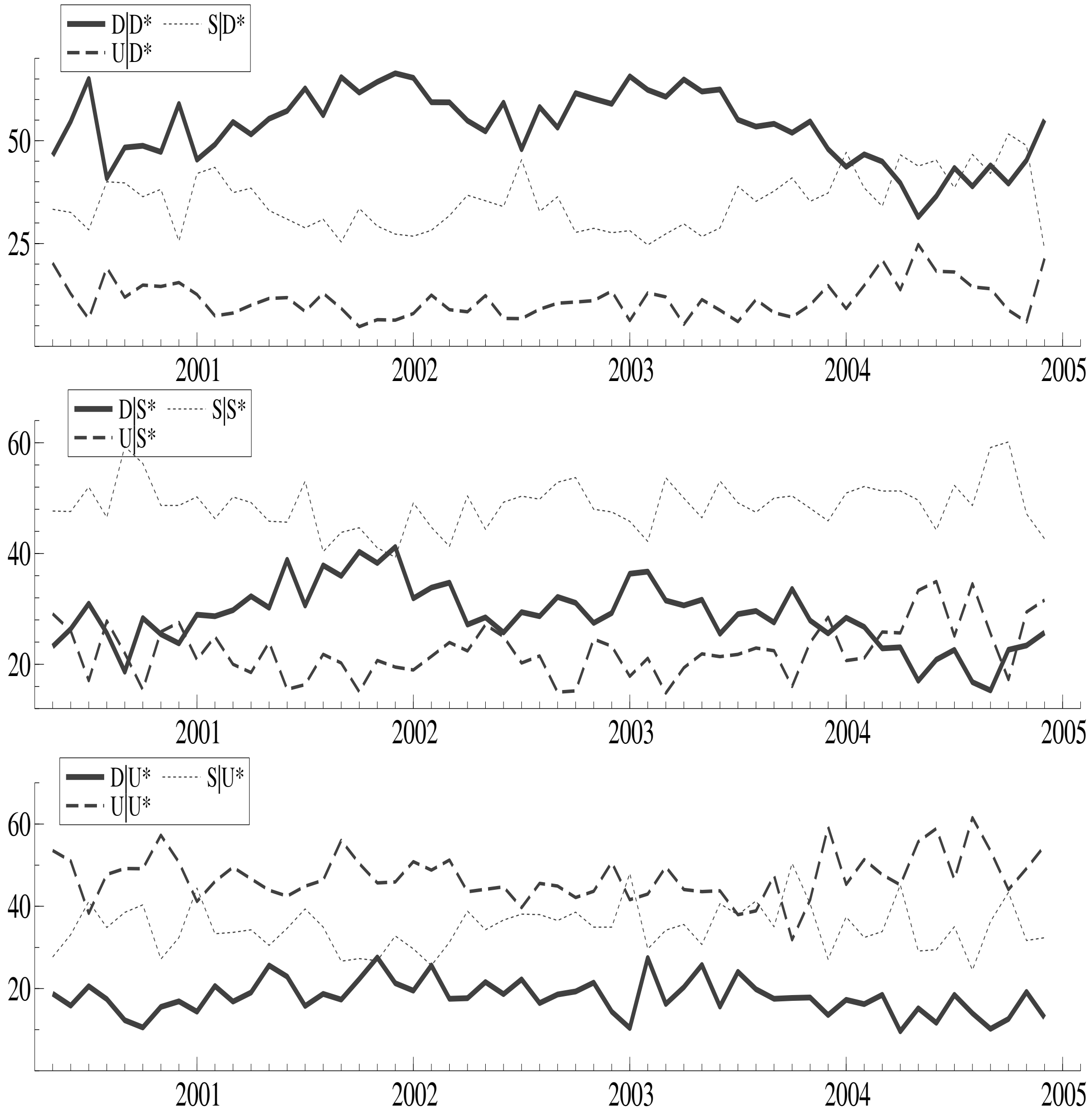

Figure 2: Estimates of $P(y^r_{it,j}|y^p_{it,k})$ (in percentages), where D, S and U denote $y^r_{it,j} = 1, 2$ and 3 respectively and D^*, S^* and U^*denote $y^p_{it,j} = 1, 2$ and 3 (down, same and up).

(U^*), no change (S^*) or a fall (D^*). It shows that the modality condition is satisfied for most months. There are 6 violations of the inequality (4), all in 2004, when firms are pessimistic. There is a just one violation for those firms that expected no change, and 4 violations for optimistic firms. We also note that only one of these violations, of the best-case scenario, is statistically significant at a 95% level and that is in 2003m10 when firms were optimistic.

Firms' expectations therefore satisfy these bounds and are consistent with the best-case scenario. But additional to Das et al. (1999), as we also see from Figure 1, this is not saying that the expectational data are necessarily that useful or reliable since, on average across time, 45% of firms who expected a 'down' did not subsequently report a 'down'; 51% of firms who expected 'no change' did not subsequently report 'no change'; and, 52% of firms who expected an 'increase' did not subsequently report an 'increase'. The best-case scenario appears quite a weak requirement; it remains important to evaluate the qualitative data against the quantitative data.

3.1.2 Median response

Let $y^p_{it,k}$ be the category that contains the α-quantile of the firm's subjective density forecast for y_{it}. Let p^*_{it} denote this $\alpha-$quantile. Therefore, when $\alpha = 0.5$ the firm reports the category that contains the median of their subjective distribution and corresponds to them minimising their absolute forecast error.

Since in the best-case scenario, y_{it} is drawn from this same subjective distribution,

$$P(y_{it} - p^*_{it} < 0) = \alpha \tag{6}$$

When the observed expectation is k ($k = 1, 2, 3$) then p^*_{it} is such that

$$a_{k-1,t} < p^*_{it} \leq a_{k,t} \tag{7}$$

implying

$$y_{it} - a_{k,t} \leq y_{it} - p^*_{it} < y_{it} - a_{k-1,t} \tag{8}$$

With (6), it then follows that

$$\begin{aligned} P(y_{it} - a_{k-1,t} &< 0 \mid y^p_{it} = k) \leq \alpha \leq P(y_{it} - a_{k,t} < 0 \mid y^p_{it} = k) && (9) \\ P(y^r_{it} &< k-1 \mid y^p_{it} = k) \leq \alpha \leq P(y^r_{it} < k \mid y^p_{it} = k) && (10) \end{aligned}$$

which implies the following inequalities for the $\alpha-$quantile category:

$$P(y_{it}^r > k \mid y_{it}^p = k) \leq 1-\alpha \tag{11}$$
$$P(y_{it}^r < k \mid y_{it}^p = k) \leq \alpha \tag{12}$$

The best-case scenario implies that for a group of firms who report $y_{it}^p = k$, the $\alpha-$quantile of the distribution of realisations fall in category k. Tests for whether (11) and (12) are satisfied for a given k and α can be then constructed using

$$\sqrt{n_{.k,t}}(\sum_{j=k+1}^{3} \widehat{p}_{jk,t} - \sum_{j=k+1}^{3} p_{jk,t}) \rightarrow N\left[0, \left(1-\sum_{j=k+1}^{3} p_{jk,t}\right)\sum_{j=k+1}^{3} p_{jk,t}\right]. \tag{13}$$

Since this test exploits the ordering of firms' responses, unlike (4), the requirements for the best-case scenario are stronger under the median response. It implies sharper bounds given that (11) implies (4) for $k = 1$ (the lowest category) and (12) implies (4) for $k = 3$ (the highest category). This means that the median category assumption requires a majority of firms to be in the expected bin when firms are either optimistic or pessimistic, while the modal category assumption requires only a plurality.

Figure 3 presents 90% confidence intervals for the probabilities in (11) and (12) for $\alpha = 0.5$. Looking first at the top panel for the pessimists we see that only for 6 months, towards the end of the sample period, do the confidence bands not contain 0.5 and we reject (11). These rejections suggest that firms may have had a lower α in mind. The middle panel of Figure 3 shows that for those firms that expected 'no change', since the bands fall below 0.5, one cannot reject (11) or (12). The bottom panel, for the optimists, shows that on 12 occasions the confidence bands are too high, now suggesting firms may have had a higher α in mind. $\alpha > 0.5$ is consistent with the view that firms are more afraid of under-predicting than over-predicting the future values of y_{it}.

Additional to the analysis of Das et al. (1999), Figure 4 provides complementary information on the balances of risks to firms. $P(y_{it}^r < k \mid y_{it}^p = k)$ is the probability that firms overpredict; $P(y_{it}^r > k \mid y_{it}^p = k)$ is the probability that firms overpredict. Under the best-case scenario, and importantly maintaining the assumption of no common/macroeconomic shocks, when $\alpha = 0.5$ we should expect an equal proportion of firms to be positively and negatively surprised. Accordingly, Figure 4 plots the proportion of firms for whom the realisation was greater than or less than expected. We see that for the majority of the sample-period firms were too optimistic. This is consistent with the view that $\alpha > 0.5$. The bottom panel of Figure 4 shows that optimism does not appear to relate to movements in aggregate IoP over the sample-period. We note that the sharp transitory movement in the IoP seen in mid 2002 was associated with seasonal adjustment difficulties in May 2002 because the Bank Holiday was moved to the first week of June and there was an additional holiday to celebrate the Queen's Jubilee; the 2002 (Football) World Cup is also believed to have distorted the typical seasonal pattern. In any case, this oddity in the IoP data does not affect our analysis below which is based on the underlying firm-level data; the aggregate data are considered only for reference purposes.

It is perfectly possible for firms' qualitative expectational data to be deemed "best-case" predictions, against the qualitative realisations data, but nevertheless not actually provide a clear signal about quantitative movements.

3.2 Mean assumption

If firms report the category that contains the mean of their subjective distribution, which corresponds to them minimising squared forecast errors:

$$E(y_{it} \mid y_{it}^p = k) \in (a_{k-1,it}, a_{k,it}] \tag{14}$$

The best-case scenario then implies that for any group of firms who expect $y_{it}^p = k$ the mean of the distribution of (quantitative) realisations falls in category k. To implement a test of (14) requires both data on the quantitative realisations y_{it} and knowledge of the thresholds $a_{k,it}$. Even when the $a_{k,it}$ are unknown, which is typically the case (e.g., in the cross-country qualitative business survey held by the European Commission the $a_{k,it}$ are determined subjectively and not reported by the survey), (14) implies that $E(y_{it} \mid y_{it}^p = k)$ should increase with k, as long as the thresholds increase in k.

3.2.1 Signal versus noise and the coherence between qualitative and quantitative data

We can also test what we call the coherence between the retrospective qualitative data and the quantitative realisations data. This is important, given that these data are from different surveys, with different sampling and data measurement assumptions. The widespread use of retrospective qualitative business surveys to provide more timely estimates of the quantitative data is implicitly predicated on the assumption that the two samples are measuring the same concept of output growth. If they are then retrospectively firms report the category that contains the quantitative realisation, drawn from their objective

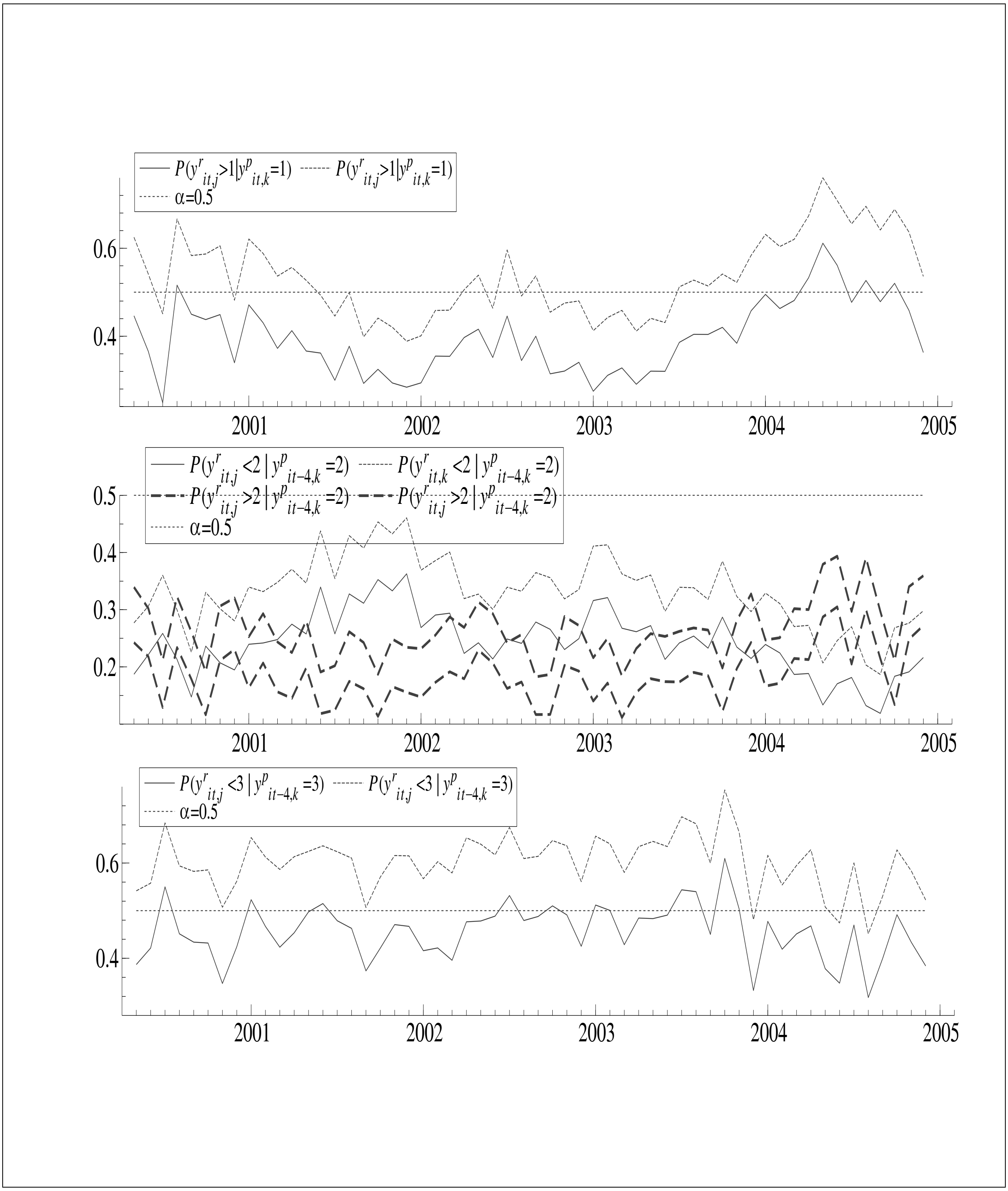

Figure 3: 90% confidence intervals for the cumulative probabilities

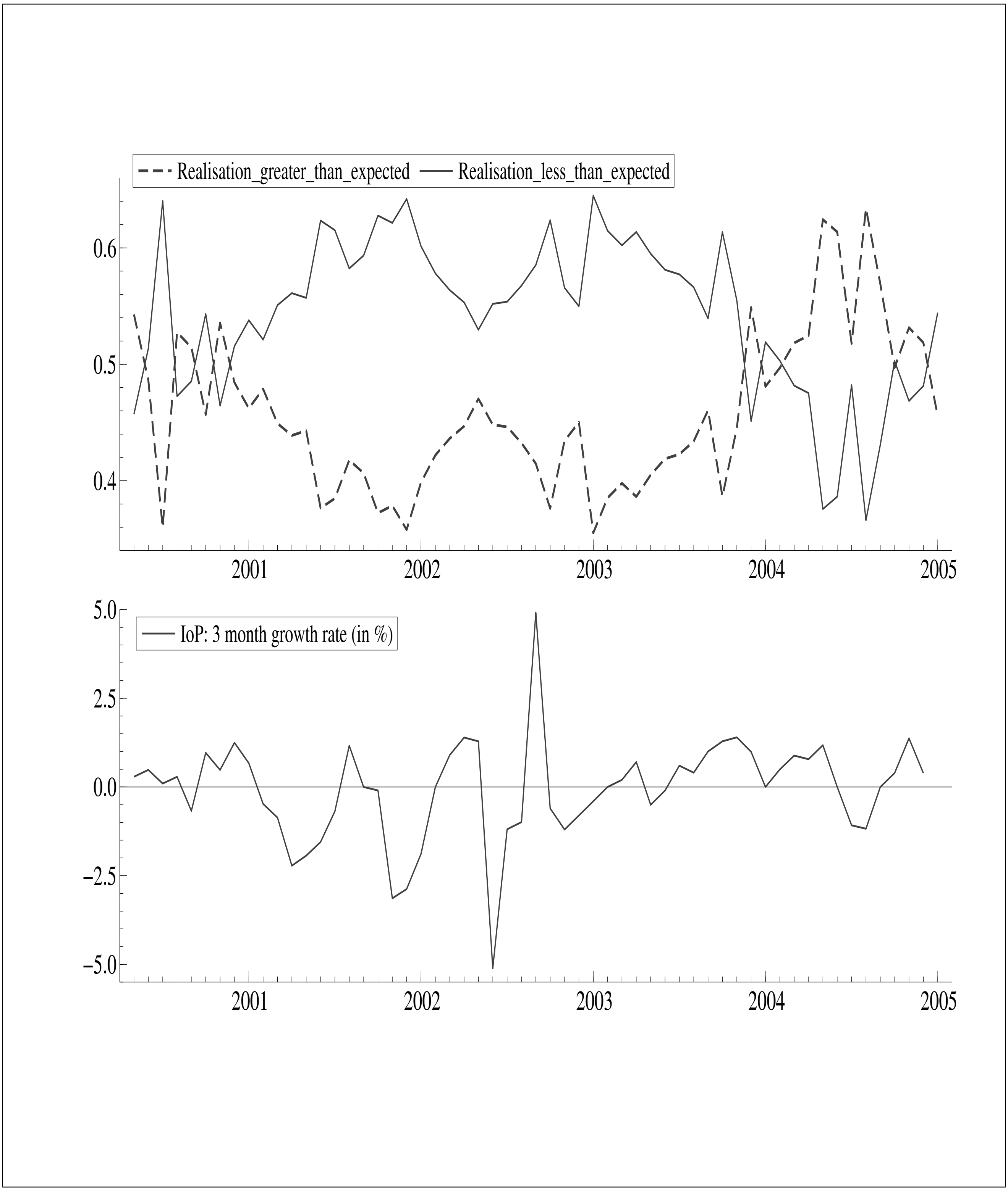

Figure 4: Probability of outturn higher than expected (upper panel) and IoP growth (lower panel)

(but unknown) density, and we should expect:

$$E(y_{it} \mid y^r_{it,k}) \in (a_{k-1,it}, a_{k,it}] \tag{15}$$

implying that $E(y_{it} \mid y^r_{it} = k)$ should also increase with k. This will hold even when the thresholds used in (15) differ from those in (14).

Under forecast rationality (and quadratic loss so that the forecaster reports the conditional mean of their forecast density), given that the realisation y_{it} equals the forecast/expectation, $E(y_{it} \mid y^p_{it} = k)$, plus an *i.i.d.* mean-zero error ε_{it} , $V(y_{it} \mid y^r_{it,k}) \geq V(y_{it} \mid y^p_{it,k})$. This inequality also explains the stylised fact (see Pesaran and Weale, 2006) that in qualitative surveys expectations, relative to realisations, are concentrated in the 'no-change' category.

Similarly, and related to (14), if $E(y_{it} \mid y^p_{it} = k)$ does not increase with k one might classify the expectational data as containing no *signal* about the subsequent quantitative realisations data: they are simply *noise*. Let $y_{jk,t}$ denote the mean (quantitative) realisation in category j, k at time t computed over $n_{jk,t}$ observations.[4]

Therefore a further test, while weaker than the best-case scenario and rational expectations, is to test whether the expectational data are "useful" and contain a(ny) signal about the quantitative realisations data. As (14) indicates, the expectational data contain a signal when

$$(y_{11,t} + y_{21,t} + y_{31,t}) \leq (y_{12,t} + y_{22,t} + y_{32,t}) \leq (y_{13,t} + y_{23,t} + y_{33,t}), \tag{16}$$

and they are simply noise when

$$H_0 : (y_{11,t} + y_{21,t} + y_{31,t}) = (y_{12,t} + y_{22,t} + y_{32,t}) = (y_{13,t} + y_{23,t} + y_{33,t}) \tag{17}$$

A t-test for (17) can be constructed for the difference between these three means, as the corresponding variance estimates can be added due to independence (the sample means for each categorical response are based on disjoint sets of observations). Rejection of the null of noise does not mean the data are rational. Rationality requires $\varepsilon_{it} = y_{it} - E(y_{it} \mid y^p_{it} = k)$ to be mean zero and serially uncorrelated, and orthogonal to information known at the time the firm formed its expectation y^p_{it}. When the error is not mean zero we should not expect $E(y_{it} \mid y^p_{it} = k) \in (a_{k-1,it}, a_{k,it}]$. However, irrespective of whether macroeconomic

[4]For notational ease we do not distinguish between sample and population moments.

shocks mean the assumptions of the best-case scenario are violated, if the qualitative data do not contain a signal about the realisations data then we should not expect them to be helpful when forecasting. Indeed, since averaging (pooling) across firms, as we do when computing $y_{jk,t}$, tends to render forecasts more robust (e.g., see Timmermann (2006) and Clark & McCracken (2009)), we should perhaps be particularly sceptical about the utility of expectational data if we cannot reject (17).

Similarly, one can test the null hypothesis that the retrospective survey data are noise:

$$(y_{11,t} + y_{12,t} + y_{13,t}) = (y_{21,t} + y_{22,t} + y_{23,t}) = (y_{31,t} + y_{32,t} + y_{33,t}) \tag{18}$$

To test whether the prospective and retrospective qualitative survey data jointly contain a(ny) signal about the subsequent quantitative realisations data one can construct tests based on individual cells in the 3×3 contingency table. For example, for the retrospective and prospective qualitative data to be *coherent* with each other and the quantitative data we should expect:

$$\begin{aligned} y_{11,t} &\leq y_{22,t} \leq y_{33,t} && (19) \\ (y_{21,t} + y_{31,t} + y_{32,t}) &\geq (y_{12,t} + y_{13,t} + y_{23,t}) && (20) \end{aligned}$$

Note $(y_{21,t} + y_{31,t} + y_{32,t})$ is a realisation greater than expected and $(y_{12,t} + y_{13,t} + y_{23,t})$ is a realisation less than expected. Under the best-case scenario we should also expect equality between the row and column sums:

$$\begin{aligned} (y_{11,t} + y_{21,t} + y_{31,t}) &= (y_{11,t} + y_{12,t} + y_{13,t}) && (21) \\ (y_{12,t} + y_{22,t} + y_{32,t}) &= (y_{21,t} + y_{22,t} + y_{23,t}) && (22) \\ (y_{13,t} + y_{23,t} + y_{33,t}) &= (y_{31,t} + y_{32,t} + y_{33,t}) && (23) \end{aligned}$$

To examine (14) and (15), and test for coherence between the two datasets, requires use of the matched dataset, introduced in Section 2. The sample-size of the ITS drops considerably when matched against the MPI and this precludes meaningful analysis across time, as above. We will therefore focus on results pooled across time and simply remark that, no doubt explained in part by the smaller sample sizes, there was considerable volatility across time in terms of the relationship between the two datasets.

Table 1 presents estimates of the sample mean of the quantitative outturns, y_{it}, pooled across firms and time, along with t-values and the number of observations, given the firms' retrospective and prospective qualitative responses, prior to Winsorisation. Table 2 then reports analogous results when y_{it} is Winsorised to mitigate the effect of outliers. The weighted (by the number of observations) sum of these sample means across the rows and columns in each table gives, respectively, the sample mean of y_{it} for each retrospective and prospective categorical response. Inspection of these indicates that only retrospectively do the sample means increase with j. Looking at the raw data, when firms report a down they contract on average by -4.0%; when they report an up they grow by 2.7%; and when they report 'no-change' one cannot reject the null hypothesis that output growth is zero given a t-value of -0.145. The t-value testing (18), using the information contained in Table 1, is -1.97, indicating that these differences are statistically significant at a 95% level. Moreover, the t-value increases to -2.76 when we simply test $(y_{11,t} + y_{12,t} + y_{13,t}) = (y_{31,t} + y_{32,t} + y_{33,t})$, indicating that optimists and pessimists certainly experience contrasting fortunes. So the retrospective survey data do contain a signal about the quantitative data. A similar picture is seen using the Winsorised data in Table 2, although the sample means do not increase so strongly with j.

But the expectational survey data do not contain a statistically significant signal, with the sample means poorly determined (with large standard errors) such that one cannot reject the null hypothesis of noise, (17), using information contained in both tables 1 and 2, with t-values of -0.11 and -0.48, respectively. Macroeconomic shocks, occurring after the forecast was made, might be thought to contribute to this finding; although they do not appear to prevent the expectational data being best-case predictions for the qualitative realisations data. But analysis of the macroeconomic data (the growth rate of the aggregate index of production, series CKYY on the ONS database, which is constructed from the firm-level MPI data we consider) reveals there to be no statistical evidence of first (or higher) order serial correlation over our sample-period. This suggests that pooling the data across time, as in Tables 1 and 2, is not unreasonable since observations appear independent (across firms) in the absence of common (macroeconomic) shocks. However, we did re-compute the tables having subtracted macroeconomic forecasts, based on recursive estimation of an autoregressive model estimated using real-time data vintages available from the ONS, and inference was qualitatively unchanged. This is unsurprising, given the absence (see Figure 4) of pronounced (aggregate) cyclical movements in this stable sample.

3.3 An apparent paradox

To summarise, our analysis of the firm-level qualitative and quantitative data has revealed that:

i.) the firm-level qualitative expectational data are best-case predictions of the outturns but, importantly, as declared qualitatively by firms;

ii.) the retrospective qualitative data and the quantitative realisations data from the two different surveys are *coherent* with each other;

iii.) the qualitative expectational data are not consistent with what we should expect if they were best-case predictions of the quantitative outturns and they do not contain a signal about the quantitative realisations data.

This is an apparent paradox. Given i.) and ii.) hold we might expect this to imply that the qualitative expectational data should be useful at explaining the quantitative realisations data, contradicting our finding in iii.). While of empirical significance, we can think of several possible explanations for this apparent paradox.

Firstly, the accumulation of 'forecasting' errors and 'discretisation' errors mean the ability of the qualitative expectational data to predict the quantitative realisations data can be drowned out by the combination of these two noise terms. This can be seen by defining the two errors as follows:

$$
\begin{aligned}
E(y_{it} \quad | \quad y_{it}^{r} = k) &= E(y_{it} \mid y_{it}^{p} = k) + u_{it} & (24)\\
y_{it} \quad &= \quad E(y_{it} \mid y_{it}^{r} = k) + u_{it}^{r} & (25)
\end{aligned}
$$

where the 'forecasting' error, u_{it}, defined in (24), is the difference between firms' retrospective qualitative assessment of their output growth and firms' qualitative forecast and reflects how firms' update their (qualitative) forecast having observed y_{it}. The 'discretisation' error, u_{it}^{r}, is the difference between the realisation, y_{it}, and firms' qualitative assessment of it. Assuming a normal distribution for y_{it} this can be written like a generalised residual (see Gourieroux et al. (1987)) so that[5]

$$u_{it}^r = y_{it} - \frac{\phi(a_{k-1,it}) - \phi(a_{k,it})}{\Phi(a_{k,it}) - \Phi(a_{k-1,it})} \tag{26}$$

where $\phi(.)$ denotes the density and $\Phi(.)$ the distribution function of the standard normal distribution. u_{it}^r reflects how much information is lost through firms reporting y_{it} qualitatively rather than quantitatively. Discretisation of (the continuous) y_{it} reduces the amount of information in the sense defined by Shannon (1948); less information would be lost if the number of states into which y_{it} was discretised was greater than three ($k = 1, 2, 3$). Substituting (24) into (25) then reveals that

$$y_{it} = E(y_{it} \mid y_{it}^p = k) + u_{it} + u_{it}^r \tag{27}$$

implying that the informational content of firms' qualitative expectations is weakened by the compounding of the two errors. Our results therefore provide firm-level support for the method introduced by Lee (1994) to account for this discretisation error when using expectational data at the aggregated/macroeconomic level. Lee (1994) finds that the evidence against the rationality of aggregated expectational data, from the CBI survey, is weaker when one conducts rationality tests based on u_{it} (aggregated across i) alone, rather than using the composite error ($u_{it} + u_{it}^r$). Similarly, we find that adding the discretisation error to the forecasting error renders the qualitative expectational data uninformative about the quantitative realisations data at the firm-level.

Secondly, firms may reply to the expectational question with the mode or median of their density forecast in mind, rather than the conditional mean $E(y_{it} \mid y_{it}^p = k)$. When they form best-case predictions in this manner i.) should hold, but the qualitative expectational data need not be best-case predictions of the quantitative outturns and need not contain a signal about them, as long as the mean forecast differs from the mode/median. But ii.) should continue to hold given that retrospectively firms do not base their qualitative response on their subjective density forecast but instead are supposedly replying by stating the category $y_{it}^r = k$ ($k = 1, 2, 3$) in which y_{it} is contained.

[5]The normality assumption is innocuous given our use of nonparametric tests. We make it here for expositional purposes only. The point we wish to make is that an error is induced, and information lost, when then the quantitative data are reduced to a qualitative response. A different distributional assumption for y_{it} would affect the form u_{it}^r takes but not eliminate it.

4 Conclusion

This paper assesses the utility of qualitative expectational survey data at the firm-level in terms of both their ability to anticipate firms' subsequent retrospective, but qualitative, reports of their performance but also these same firms' quantitative answers. The former assessment uses nonparametric tests of the "best-case scenario", developed by Das et al. (1999), that identify bounds on the distribution of realised outcomes conditional on firms' qualitative expectations under the assumption of rational expectations. The latter assessment requires access to a unique panel dataset which matches firms' responses to a leading qualitative tendency survey conducted by the Confederation of British Industry with these same firms' quantitative replies to a different survey carried out by the Office for National Statistics. We introduce a weaker test for the *coherence* between these two surveys and test whether the qualitative data contain a(ny) signal about the quantitative data. This lets us test whether the two samples are measuring the same concept of output growth as implicitly assumed when qualitative business surveys are used to provide forecasts, and more timely estimates, of the quantitative data.

We find that while firms' qualitative expectations of their future output growth are best-case predictions of their retrospective qualitative assessment of this growth, they do not contain a signal about the quantitative data. But the retrospective qualitative data are coherent with the quantitative data, suggesting that the two surveys are measuring similar concepts despite different questions and different sampling assumptions. We explain that the apparent paradox that the qualitative expectational data do not help predict the quantitative realisations data is explained by 'forecasting' and 'discretisation' errors confounding any signal from the expectational data. Moreover, when reporting their expectations qualitatively firms may report the mode or median of their subjective density forecast rather than its mean. When there are pronounced asymmetries in this density forecast (and the mean differs from the mode and/or median) the qualitative expectational data can be best-case predictions of the qualitative but not the quantitative realisations data.

These results suggest that qualitative business survey data, given that they are published ahead of the ONS's quantitative data, are likely to prove more useful for nowcasting than forecasting. While aggregation of the firm-level expectational data, if firms' forecasting and disretisation errors are offsetting, may improve the informational content of macroeconomic indicators constructed from these expectational data, it would have been reassuring for those using the aggregated expectational data if any macroeconomic indicator was underpinned by evidence that the firm-level expectational data were not simply noise (i.e., if (17) had been rejected statistically).

Another implication, given a demand for useful expectational data, is that it would help if the business surveys elicited quantitative rather than qualitative responses to the expectational question. Information is lost when firms discretise their responses; our application shows that the extra noise induced by discretisation contributes to the expectational data being unable to predict the quantitative realisations data, even though they are able to predict firms' qualitative assessments of these realisations data. Our results provide firm-level support for the method introduced by Lee (1994) to account for this discretisation error when using expectational data at the aggregated/macroeconomic level.

We note that in principle one can test the utility of expectational data at the firm-level for other countries using the methods we set out. This may require the permission of the survey producers. As in the UK, the qualitative and quantitative surveys are often carried out by different institutions. But further applications would be helpful in establishing a consensus at the underlying firm-level on the utility, or otherwise, of expectational data to economic forecasters.

Table 1: Sample mean of the quantitative outturns (raw data), y_{it}, pooled across firms and time, given the firms' retrospective and prospective qualitative responses. t-values in (.), and the number of observations in {}.

	j/k	$y^p_{it,k}$ 1	2	3	Row Sum
	1	-2.689	-4.790	-5.401	-4.028
		(-1.268)	(-2.757)	(-2.157)	(-3.312)
		{528}	{586}	{190}	{1304}
$y^r_{it,j}$	2	2.304	-0.946	0.091	-0.132
		(0.976)	(-0.836)	(0.045)	(-0.145)
		{387}	{1274}	{417}	{2078}
	3	-4.087	2.414	4.444	2.738
		(-1.127)	(1.314)	(2.389)	(2.217)
		{127}	{549}	{612}	{1288}
Column Sum		-1.005	-1.115	1.420	
		(-0.690)	(-1.320)	(1.155)	
		{1042}	{2409}	{1219}	

Notes: t-tests for (17) and (18) can be constructed from the information in these tables since the variance estimates, underlying the t-values in (.), can be added due to independence (the sample means for each categorical response are based on disjoint sets of observations)). The t-value for (17) is -0.11 and the t-value for (18) is -1.97.

Table 2: Sample mean of the quantitative outturns (Winsorised data), y_{it}, pooled across firms and time, given the firms' retrospective and prospective qualitative responses. t-values in (.), and the number of observations in {}.

	j/k	$y^p_{it,k}$ 1	2	3	Row Sum
$y^r_{it,j}$	1	-2.211 (-1.729) {528}	-3.239 (-2.807) {586}	-5.298 (-2.652) {190}	-3.123 (-3.961) {1304}
	2	1.608 (1.158) {387}	-0.275 (-0.373) {1274}	0.135 (0.103) {417}	0.158 (0.271) {2078}
	3	-1.224 (-0.450) {127}	1.016 (0.848) {549}	2.232 (1.905) {612}	1.373 (1.712) {1288}
Column Sum		-0.672 (-0.753) {1042}	-0.701 (-1.268) {2409}	0.341 (0.424) {1219}	

Notes: See notes to Table 1. The t-value for (17) is -0.48 and the t-value for (18) is -2.14

5 Appendix

5.1 The MPI

The MPI adopts stratified random sampling, stratifying the population by industry and employment. There are four employment sizebands within each industry but the numbering of these is dependent upon the industry cut-off. This is a level above which all contributors within that band are required to respond, that is, they do not sample. The decision on a cut-off is based on the number of contributors within each industry. Samples for the MPI are derived from the IDBR via a Permanent Random Number system, which allows gradual rotation of the sample within each stratum for each four-digit industry. Sampled firms stay in the sample for a period of months and then receive a "holiday period" which allows them to stay out of the sample. However, those firms in the employment sizeband for which the sampling fraction equals 100% stay in the sample permanently. These are usually large firms.

There are about 160,000 businesses in the sector covered by the MPI and the overall sample size in each month is about 9000, of which 600 receive employment only forms. The sampling unit of the MPI survey is the reporting unit. The reporting unit of a firm holds the mailing address to which the form is sent. The form can thus cover the whole enterprise, or parts of the enterprise identified by lists of local units.

5.2 The ITS

The ITS was originally carried out on a quarterly basis. A monthly inquiry has also been carried out in intervening months using five questions selected from the quarterly survey form. The participating UK manufacturing firms are required to give qualitative information about their past and future expectations on their volume of output, costs, prices, business confidence, employment, and some other related questions. Firms are removed from the sample i) if they go out of business or ii) they decide not to participate further. The sample size has gradually decreased over time. The majority of the ITS survey forms are sent to the headquarters or parent companies who generally respond on behalf of their UK-based activities. In the cases of small or medium enterprises, forms are sent to their sole addresses.

References

Ashley, J., Driver, R., Hayes, S. & Jeffery, C. (2005), 'Dealing with data uncertainty', *Bank of England Quarterly Bulletin* **Spring**, 23–29.

Clark, T. E. & McCracken, M. W. (2009), 'Averaging forecasts from VARs with uncertain instabilities', *Journal of Applied Econometrics* . Forthcoming. Revision of Federal Reserve Bank of Kansas City Working Paper 06-12.

Claveria, O., Pons, E. & Ramos, R. (2007), 'Business and consumer expectations and macroeconomic forecasts', *International Journal of Forecasting* **23**(1), 47–69.

Das, M., Dominitz, J. & van Soest, A. (1999), 'Comparing predictions and outcomes: Theory and application to income changes', *Journal of the American Statistical Association* **94**, 75–85.

Dixon, W. (1960), 'Simplified estimation from censored normal samples', *Annals of Mathematical Statistics* **31**, 385–391.

Driver, C. & Urga, G. (2004), 'Transforming qualitative survey data: performance comparisons for the UK', *Oxford Bulletin of Economics and Statistics* **66**, 71–90.

Gourieroux, C., Monfort, A., Renault, E. & Trognon, A. (1987), 'Generalized residuals', *Journal of Econometrics* **34**, 5–32.

Gourieroux, C. & Pradel, J. (1986), 'Direct test of the rational expectation hypothesis', *European Economic Review* **30**(2), 265–284.

Horvath, B., Nerlove, M. & Wilson, D. (1992), A re-interpretation of direct tests of forecast rationality using business survey data, *in* K. H. Oppenlander & G. Poser, eds, 'Business Cycle Analysis by Means of Economic Surveys, Part 1', Avebury, Aldershot.

Ivaldi, M. (1992), 'Survey evidence on the rationality of expectations', *Journal of Applied Econometrics* **7**(3), 225–41.

Lee, K. (1994), 'Formation of Price and Cost Inflation Expectations in British Manufacturing: a Multisectoral Anlaysis', *Economic Journal* **104**, 372–386.

Lui, S., Mitchell, J. & Weale, M. (2008), Qualitative business surveys: Signal or noise? National Institute of Economic and Social Research Discussion Paper No.323.

Manski, C. (1990), 'The use of intentions data to predict behavior: A best case analysis', *Journal of the American Statistical Association* **85**, 934–940.

Mitchell, J., Smith, R. & Weale, M. (2005), 'Forecasting manufacturing output growth using firm-level survey data', *The Manchester School* **73**, 479–499.

Nerlove, M. (1983), 'Expectations, plans, and realizations in theory and practice', *Econometrica* **51**(5), 1251–79.

ONS (2005), Report on the full triennial review of the monthly production inquiry. Available at http://www.statistics.gov.uk/downloads/reviews/MPItriennialReport2005.pdf.

Pesaran, M. H. & Weale, M. R. (2006), Survey Expectations, *in* G. Elliott, C. W. J. Granger & A. Timmermann, eds, 'Handbook of Economic Forecasting Volume 1', North-Holland, pp. 715–776.

Shannon, C. (1948), 'A mathematical theory of communication', *The Bell Systems Technical Journal* **27**, 379–423, 623–656.

Timmermann, A. (2006), Forecast combinations, *in* G. Elliott, C. W. J. Granger & A. Timmermann, eds, 'Handbook of Economic Forecasting Volume 1', North-Holland, pp. 135–196.