Qualitative Business Surveys: Signal or Noise?

Silvia Lui, James Mitchell and Martin Weale*
National Institute of Economic and Social Research

November 18, 2008

Abstract

This papers identifies the information content at the firm-level of qualitative business survey data by first examining the consistency between these data and the quantitative data provided by the same respondents to the UK's ONS in official surveys. Since the qualitative data are published ahead of the quantitative data the paper then assesses the ability of the qualitative data to predict (or nowcast) the firm-level quantitative data.

Keywords: Early indicators; Firm-level comparison; Information content; Matched dataset; Qualitative business survey data; Quantitative official survey data

J.E.L. Classification: C81, C53, C23, E27

---

*Address for correspondence: Silvia Lui, National Institute of Economic and Social Research, 2 Dean Trench Street, Smith Square, London, SW1P 3HE, U.K. Tel: +44 (0) 207 654 1927. Fax: +44 (0) 207 654 1900. E-Mail: 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 Rhys Davies (ONS), Robert Gilhooly (ONS), Felix Ritchie (ONS), Richard Welpton (ONS) and Jonathan Wood (CBI). We also thank the members of the project board, including representatives of the Bank of England, CBI and ONS, convened to advise this project, for helpful comments. Thanks also to participants in presentations at the National Statistics Methodology Advisory Committee (NSMAC) and the 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.

1 Introduction

The purpose of this paper is to assess the relationship, at the firm-level, between qualitative business survey data produced by the Confederation of British Industry (CBI) and quantitative data collected by the Office for National Statistics (ONS) in the UK which serve as the basis for construction of the national accounts. Interest focuses on the CBI survey and similar surveys because they are more timely than ONS data, although the cost of that timeliness is their qualitative nature. Firms are asked a range of questions to which they provide categorical instead of quantitative answers; for example they are asked whether output has fallen, stayed the same or risen but not by how much it has changed. Similar qualitative surveys exist in many other countries; indeed the CBI survey is the basis for the UK data maintained in the European Commission's Euro-wide database of business and consumer surveys.

In recent years doubts about official data have resulted in increased prominence being attached to qualitative survey data of this type (see Ashley et al. (2005)). However past studies of the relationship with official data have largely relied on comparisons between summary statistics from the qualitative survey (such as the extent to which the proportion of firms reporting a rise in activity exceeds the proportion reporting a decline) and aggregate data from the national statistical office, such as the percentage change in output; see Driver & Urga (2004) and, for a survey, Pesaran & Weale (2006). At best, there have been comparisons between aggregate ONS data and the individual responses collected by the CBI but these too provide only a limited picture of the relationship between the two sets of data (see Mitchell et al. (2005, 2006)). Thus, on the basis of these studies it is difficult to say how firmly perceived relationships between the two data sets are based.

Therefore, in this paper, we compare the individual responses provided to the CBI with those collected by the ONS on a firm-by-firm basis and test whether a relationship exists and, if so, identify its form. This represents an entirely new approach to provide a definitive answer to the question, how useful are qualitative surveys like the CBI's? For the qualitative survey data to be a reliable indicator at the macroeconomic level it would be reassuring to find that they are related to official data at the firm-level. Since the qualitative survey asks firms about what has happened in recent months, and its link with the quantitative data is likely to depend on the period of the quantitative data under consideration, we identify when in the past the signal from the firm-level qualitative survey was strongest, and when, if at all, it was just noise. Finally to quantify the *value* to economists of any relationship between the two surveys we develop a means to assess the ability of the qualitative data to predict (or nowcast) the firm-level quantitative data.

To match firms' responses across the two surveys we arranged for the CBI to provide their data set to the ONS in a manner which preserves obligations of confidentiality for both bodies. The ONS then matched the CBI data set to its own Monthly Production Inquiry (MPI) for the five years 2000-2004, inclusive. This allowed us to match up the response provided by each firm in answering the CBI survey with the response provided by the same firm in providing quantitative information to the ONS in its MPI. 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.

The plan of the remainder of the paper is as follows. Section 2 describes the matched panel dataset and discusses some statistical issues. Section 3 provides some descriptive evidence summarising the relationship between firms' qualitative and quantitative responses, and also contrasts it with the relationship found at the macro-economic level. Section 4 then supplements this with more formal econometric modelling which lets us both test whether the qualitative survey are noise, and helps us identify where the signal is strongest. Importantly, the models let us examine the dynamic relationship between the qualitative and quantitative data. A firm-level indicator of output growth used to evaluate the predictive power of the qualitative survey data is then introduced. Section 5 presents the econometric results and Section 6 concludes.

2 The Matched Firm-level MPI and ITS Dataset

The two underlying surveys of interest are the ONS's MPI and the CBI's Industrial Trends Survey (ITS). In the MPI the ONS asks close to 9000 firms each month for quantitative information on their turnover values in the month; see Table 1 for summary information on both the MPI and ITS. (Appendix 1 provides additional details.) The MPI questionnaires are sent out to firms three days before the end of a calendar month. The aggregated (across firms) responses of the MPI are then used to construct the monthly IoP.[1]

The ITS asks firms many questions, only some of which are 'verifiable' (i.e. testable against official data). In this paper we focus on the retrospective question which currently takes the form, "Excluding seasonal variations, what has been the trend over the past three months with regard to volume of output?". Firms reply "up", "same" or "down"

[1]The collection of MPI data stops on the 18th working day of a calendar month after the survey. Data are then passed to the IoP team on the 19th working day. Responses after then can also be taken on. The IoP is published on the 26th working day after the end of the survey month.

(and “not applicable”). Until June 2003, firms were asked for their trend over the last four months.[2] The ITS is published at the end of the month concerned, and is therefore published ahead of the MPI explaining its potential value as a timely source of information about the current state of the economy.

Other questions in the ITS (e.g. “uncertainty about demand”) cannot be verified. Nevertheless, we might deem it encouraging if we found that there were a strong signal from those questions which could be verifiable against “official” (ONS) data.

Information is neither available on precisely who, at a given firm, in fact fills out the survey form nor on whether at a given firm this person varies 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.

The matched ITS/MPI dataset, the focus of our statistical and econometric analysis, was identified by first matching the panel of firms which replied to the ITS against the ONS’s Interdepartmental Business Register (IDBR). The IDBR is a list of UK businesses accounting for almost 99% coverage of economic activity. As column B in Table 1 shows a few more than 800 firms, on average, reply to the ITS each month, although there is some variation over time which we capture in Table 1 by reporting yearly averages. Over the five years 2584 firms were sampled at least once by the CBI.[3]

Text matching based on common variables, namely the names and addresses of firms, was used to match these CBI firms against those firms on the ONS’s IDBR. Of the 2585 firms who replied at least once to the ITS 2120 (82%) were successfully matched against the IDBR; see column C in Table 1. Of these 2120 firms, as column D indicates, there was a definite match (defined as when the ONS is at least 80% confident in the match) between firms across the two surveys for 1895 firms, i.e. for about 90% of firms. For the remaining firms, amounting for the difference between columns C and D in Table 1, there was a non-unique mapping between their ITS and IDBR reference numbers. This obviously creates confusion about whether a given firm in the IDBR is the same firm as in the ITS; on the advice of the ONS these firms were therefore dropped from our analysis to minimise the risk of matching errors.

The subset of ITS firms on the IDBR which are sampled by the MPI could then be extracted. Column E in Table 1 shows that, on average (across the five years), there are

[2]Our sample covers the period January 2000 to December 2004, i.e. 60 months. The trend period for the ITS retrospective output question, therefore, is four months in the first 42 months of our sampling period, but three months in the last 28 months.

[3]The identity of a small number of firms (25 firms) is anonymous; these firms are removed from our dataset.

Table 1: Average number of firms for each year in the raw datasets and the matched response dataset

Year	A MPI firms	B ITS firms	C* ITS firms registered on IDBR	D ITS firms registered on IDBR with unique ITS/MPI ref no	E Matched firms
2000	8993	740	607	534	159
2001	8966	848	696	618	173
2002	8897	853	699	631	179
2003	8811	860	705	647	175
2004	8793	787	646	597	169

* Notes: The match rate of the ITS against the IDBR is 82 percent over the 5 years - the figures in column C are computed as 82 percent of those in column B

about 170 firms present in the matched dataset each month. This represents an average (in a year) match rate against the ITS of about 21% of firms (column E divided by column B). Overall, since 2584 firms were sampled at least once by the CBI the match rate is 31%.

It would be strange if from two "random" samples (i.e. the MPI and ITS) the matched dataset picked up firms that were either particularly *good* or *bad* at replying to the ITS and we therefore proceed on the assumption that sample selection does not affect our matched dataset.

In total, across the five years, 807 firms gave at least one contemporaneous response to both the ITS and MPI. Given that about 170 firms are present in the matched dataset each month this indicates that there is considerable attrition in the matched dataset, reflecting both sample rotation particularly of small firms (see Appendix 1 for more on the sampling design of the MPI and ITS) and a firm not replying to one or both of the surveys. Indeed 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.

2.1 Statistical issues

Before comparison of the two surveys can be carried out a range of statistical issues need to be addressed. The MPI turnover data are first deflated, using the IoP deflator at the 4 digit level, to produce figures in volume terms. The MPI data are not seasonally adjusted while respondents to the ITS report "after taking seasonal effects into account". In the absence of seasonally adjusted data from the ONS at the firm or sectoral level, we prefer to incorporate any seasonal effects into the models we use to assess the relationship between the MPI and ITS data. Accordingly, seasonal dummies are used in the econometric analysis. Moreover, since the ITS elicits a response about a growth rate rather than a level, as in the MPI, the MPI data are differenced prior to analysis; amongst various filters we do consider the 12-month difference (a seasonal filter).

Specifically, let x_{it} denote the volume of turnover for firm i for time (month) t. The k-month growth rate is defined as

$$z_{it}^{k} = \ln x_{it} - \ln x_{it-k}, \ (k = 1, ..., 12), \tag{1}$$

while the rolling 1-month growth rate, which we also consider to facilitate interpretation, is defined as

$$\Delta x_{it-k} = \ln x_{it-k} - \ln x_{it-k-1} = z_{it-k}^{1}, \ (k = 0, ..., 11). \tag{2}$$

In addition, the turnover growth rates for each firm are standardised (at a firm-level) prior to analysis to mitigate the possible effects of outliers. In fact results below, which are presented unless stated otherwise using the standardised data, are very similar when the raw data are used; in general, as expected, estimates are a little sharper with the standardised data.

3 Descriptive statistics

Table 2 shows, pooled across firms and time, the average (quantitative) turnover growth rate in both the standardised and non-standardised MPI datasets, conditional on firms' qualitative answers to the ITS. Results are shown both for the full matched panel set and a sub-sample of 96 firms from the full sample, which we introduce below. In both cases, we see that when relating the ITS to turnover growth over the last month, i.e. Δx_{it}, there was not a tendency for more optimistic firms (i.e. firms who replied to the ITS saying their turnover had gone up) to have experienced greater turnover growth. For example, over the full matched non-standardised sample, pessimistic firms, in reality, saw their turnover rise by 0.53% while optimistic firms saw their turnover fall, but only slightly, by -0.08%. However, when relating firms' qualitative answers to Δx_{it-1} we do see that optimistic firms experienced greater growth than pessimistic firms, although only at Δx_{it-2} do we also see those firms who reported their output growth had not changed experience output growth in between the optimists and pessimists. In addition, at Δx_{it-2} we observe larger differences between the turnover growth rates of optimistic and pessimistic firms.

Table 2: Mean growth rate (×100) for MPI turnover given firms' qualitative responses

Non-standardised (full matched panel)									
	Δx_{it}			Δx_{it-1}			Δx_{it-2}		
	down	same	up	down	same	up	down	same	up
Mean	0.53	0.65	-0.08	0.27	0.19	0.72	-1.71	0.19	3.11
NT	1977	2948	1862	1464	2246	1389	1388	2160	1352
Standardised (full matched panel)									
	Δx_{it}			Δx_{it-1}			Δx_{it-2}		
	down	same	up	down	same	up	down	same	up
Mean	1.68	0.10	-1.64	-1.99	0.84	0.76	-6.38	-1.04	8.38
NT	1977	2948	1862	1464	2246	1389	1388	2160	1352
Non-standardised (sub-panel of 96 firms)									
	Δx_{it}			Δx_{it-1}			Δx_{it-2}		
	down	same	up	down	same	up	down	same	up
Mean	2.33	-0.29	-0.06	-0.61	-0.01	0.89	-4.08	-0.08	4.12
NT	264	469	300	264	469	300	264	469	300
Standardised (sub-panel of 96 firms)									
	Δx_{it}			Δx_{it-1}			Δx_{it-2}		
Mean	4.22	1.81	-2.98	-5.47	2.86	-2.80	-7.63	-4.58	10.69
NT	264	469	300	264	469	300	264	469	300

Figure 1 then reports the pooled (across firms and time) polyserial correlation, and 95% confidence intervals, between the ITS data and both the k-month and rolling 1-month MPI turnover growth rates. A histogram is also presented indicating the number of observations, pooled across firms and time, used to compute the correlation coefficient for a given value of k.

The polyserial correlation estimator, developed by Olsson et al. (1982), is appropriate when a continuous variable (MPI turnover growth) is correlated with a dichotomous or ordinal variable, which is assumed to be triggered by an underlying latent variable (ITS output trend) as it crosses thresholds. The estimator is computed via maximum likelihood on the assumption that the underlying continuous variables follow a bivariate normal distribution. The standard error of the correlation coefficient is extracted from the Hessian evaluated at the maximum. The polyserial correlation can be interpreted like Pearson's correlation coefficient; a large positive value indicates high positive correlation between the ITS and MPI data. But, because it is unclear both what constitutes a *large* correlation and whether this translates into improved inference about the MPI data, in Sections 4.2 and 5.2 below we assess explicitly the *value* to economists of any information contained in the ITS.

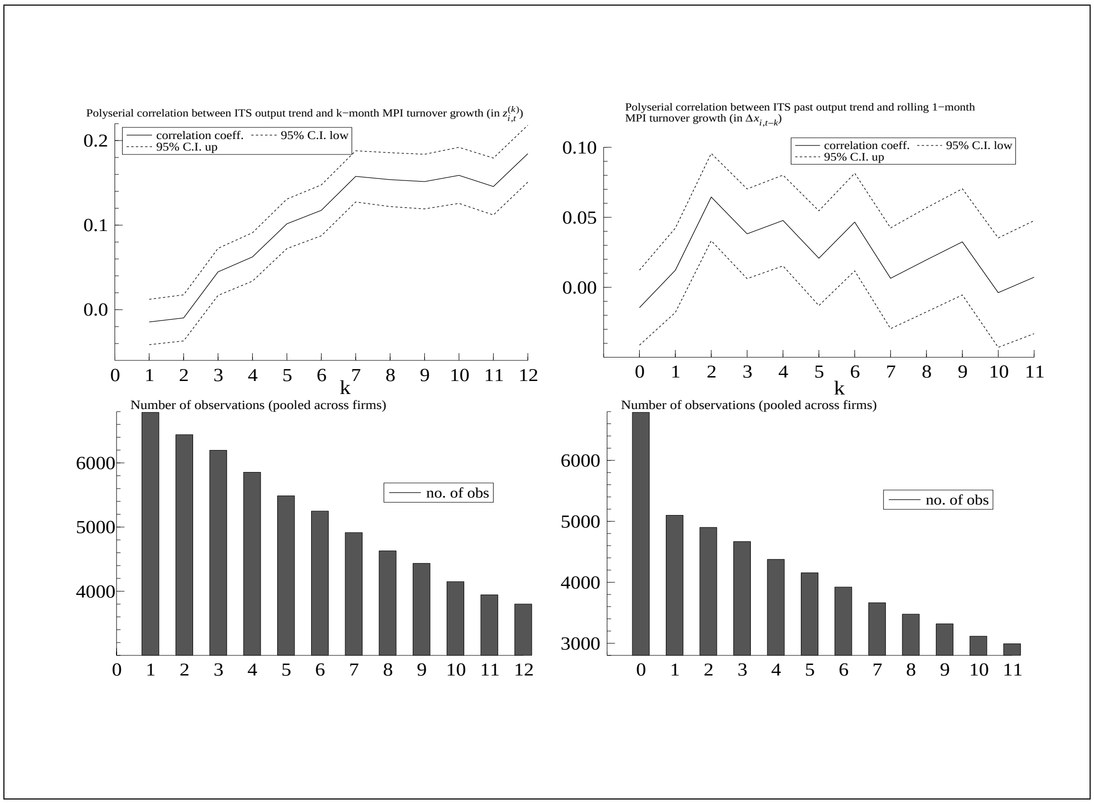

Figure 1: Pooled polyserial correlation coefficient between the firm-level qualitative and quantitative survey data, plus a histogram indicating the sample size

Figure 1 indicates that the correlation between the firm-level ITS and MPI data, z_{it}^{k}, increases with k, meaning the qualitative data are indeed a lagging indicator. This is to be expected, since the ITS question asks about growth over the last three/four months, although it is unclear how exactly firms interpret this. The right-hand-side of Figure 1 sheds more light on what the respondents to the ITS actually had in mind. It indicates that the correlation against z_{it}^k increases in k but at a decreasing rate. Specifically, the correlation of the qualitative data against Δx_{it-k} only becomes statistically significant, at a 95% level of significance, when $k \geq 2$. In other words, the firm-level ITS data at time t do not contain a statistically significant signal about growth in the last month, but about growth two months ago relative to three months ago. As k increases, the correlation against Δx_{it-k} decreases, and the signal is statistically significant (at 95%) for growth up to five, relative to six, months ago.

Interestingly, Figure 2 shows that a similar relationship holds at the macroeconomic level (i.e., for the manufacturing sector as a whole). The figure plots the correlation between the past output trend "balance statistic" from the ITS and ONS (seasonally adjusted) data on manufacturing output growth, Δx_{t-k} (defined analogously to (2)), from 1985m1-2007m12. The balance statistic is defined as the percentage of firms who report output as having risen less the percentage who report it having fallen.

As with Figure 1, Figure 2 plots the correlation against the k-month growth rate, in the top panel, and the rolling 1-month growth rate, in the bottom panel. Looking at the top panel, as with Figure 1, it is seen that the correlation with what has happened in the past month or two months is small, but that the correlation rises steadily with the length of the interval considered. The bottom panel of Figure 2 shows that the peak signal about growth is six to seven months ago. The peak correlation with the firm-level data was a little earlier, at two to three months. But both the firm-level and macroeconomic evidence suggest that the qualitative survey data do not provide a good coincident indicator of growth. The firm-level correlations suggest that when answering a question about what has happened "over the previous three to four month", firms are discussing what happened over, at least, the past six months. Thereafter, the signal is not statistically significant at 95%. By slight contrast, perhaps, the macroeconomic correlations suggest that firms look back slightly further into the past. However, these relationships are partial and ignore the likely dynamic behaviour of these data. The econometric analysis which follows addresses this.

Exploiting information from the ITS about the size and sector of each firm, Figures 3 and 4 present the polyserial correlations for selected "types" of firm. This provides an indication of whether firms are heterogeneous, in terms of the informational content of their qualitative replies to the ITS. Figure 3 presents the correlation coefficients for five sub-sectors of manufacturing, while Figure 4 presents them for small, medium and large-sized firms. Small firms are defined as those with fewer than 199 employees, medium-sized firms as those with between 200 and 499 employees and large-sized firms as those with

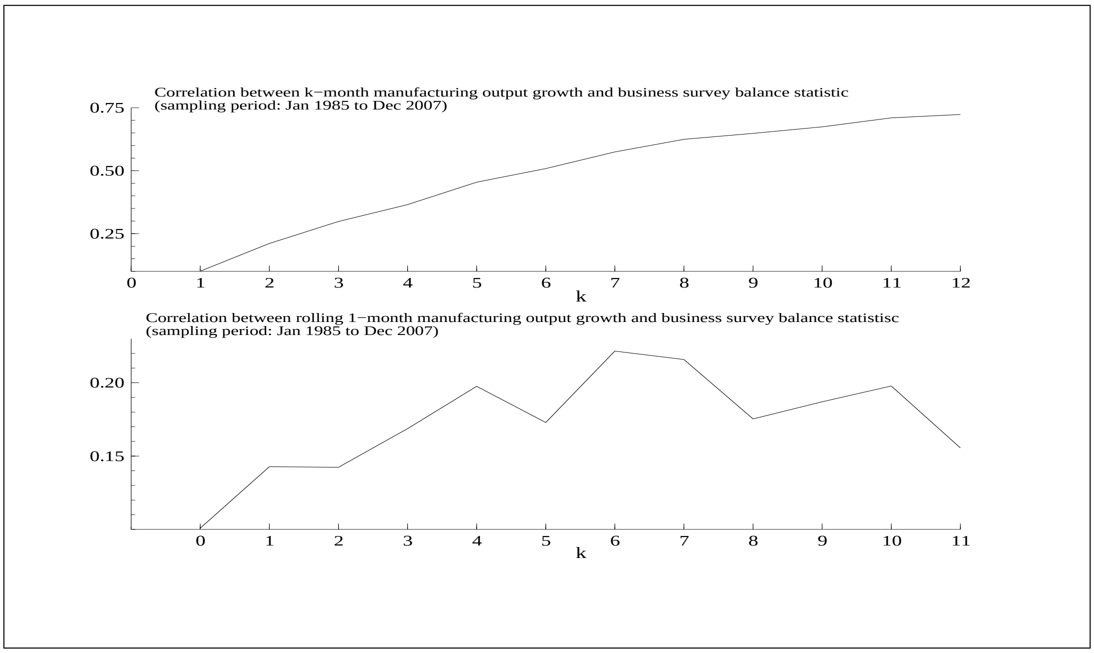

Figure 2: Correlation between the ITS and MPI data at the macroeconomic level

more than 500 employees. Both Figures indicate, despite some volatility explained in part by fewer observations than in Figure 1, that there is no obvious pattern according to the type of firm. The qualitative replies of firms from different sectors, and of different sizes, display a similar correlation against the quantitative data; in general, the point polyserial estimates for firms of different size and sector fall within the 95% confidence intervals presented around the pooled correlation estimates.

4 Assessment of the Reliability of the ITS Data

To assess formally the reliability of the CBI data we model the relationship between the ITS and MPI data, allowing for their dynamics, and test various hypotheses about the former.

The ITS asks each firm, i ($i = 1, \cdots, N_t$) at time t ($t = 1, \cdots, T_i$), to give qualitative answers to the question about its trend of output (excluding seasonal variations) over the past three/four months. As discussed, the firm can respond either 'up', 'same' or 'down', denoted as j (where $j = 2, 1, 0$, respectively).[4] Assume that there is a continuous latent variable y_{it}^* that triggers firm i's categorical response at time t via the following

[4] Firms can, and a very small number do, also respond 'not applicable'. We ignore these firms below.

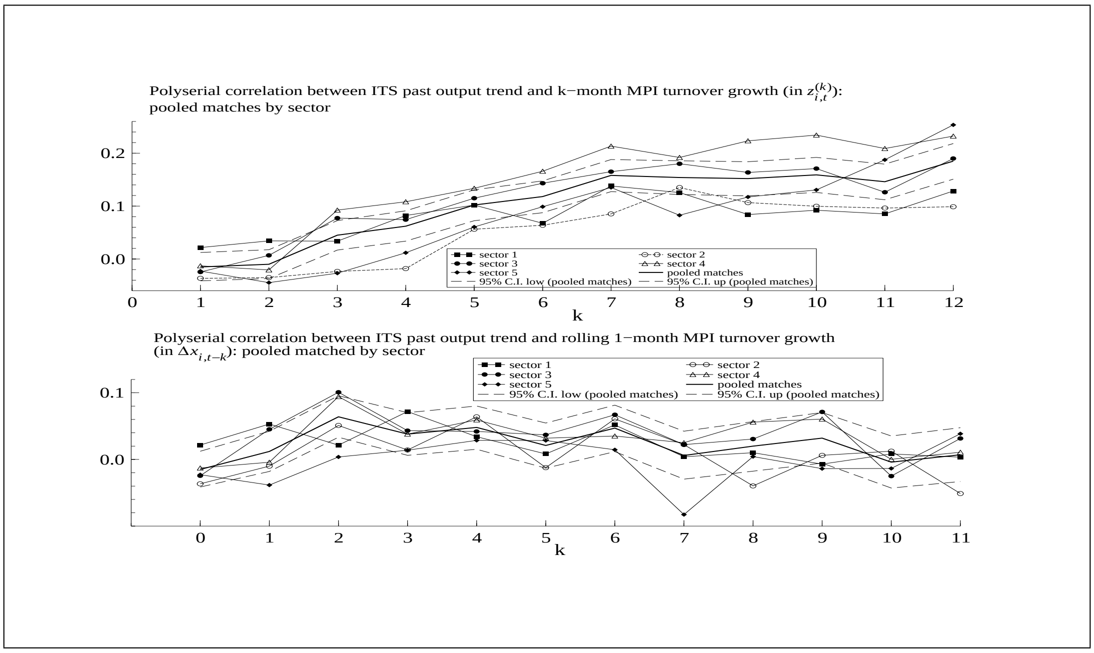

Figure 3: Polyserial correlation coefficient between the qualitative and quantitative data at the sectoral-level

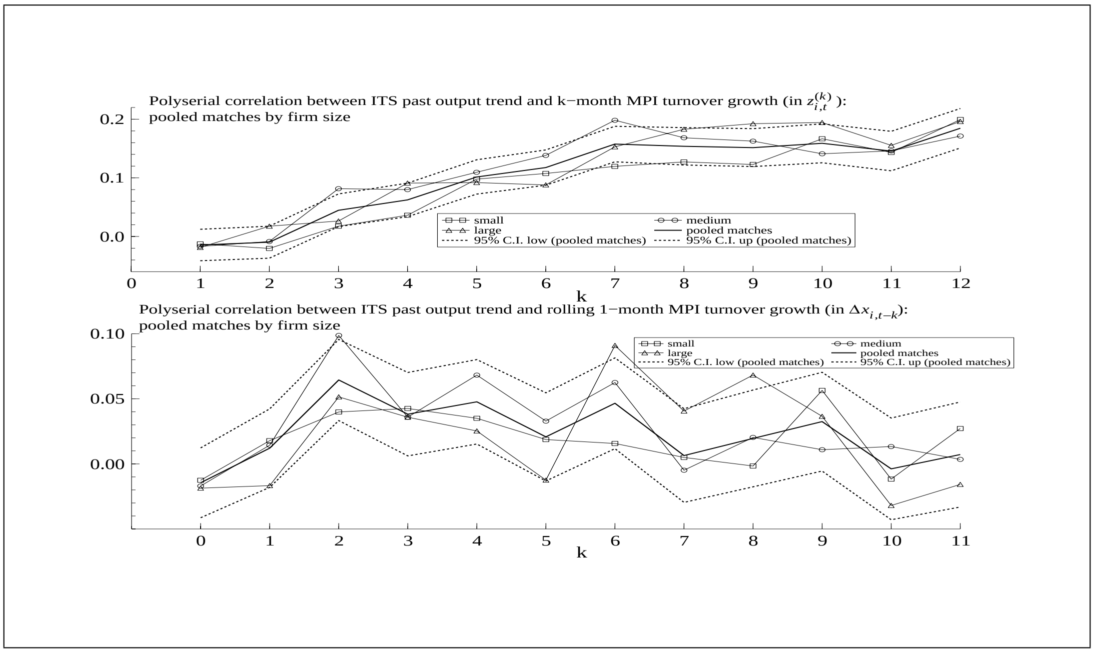

Figure 4: Polyserial correlation coefficient between the qualitative and quantitative data for small, medium and large firms

observation rule:

$$y_{it} = j \text{ if } \mu_j < y^*_{it} \leq \mu_{j+1},\ j = 0, 1, 2, \tag{3}$$

where μ_j's are the unknown thresholds: $\mu_0 = -\infty$, $\mu_j \leq \mu_{j+1}$ and $\mu_3 = \infty$.

Our starting point is then to assume that y^*_{it} provide noisy estimates of the quantitative data, $z^{(k)}_{i,t}$, whereby we consider the following general model:

$$y^*_{i,t} = \beta_1 z^{(1)}_{i,t} + \beta_2 z^{(2)}_{i,t} + \cdots + \beta_{12} z^{(12)}_{i,t} + \lambda_1 y_{i,t-1} + ... + \lambda_{12} y_{i,t-12} + \alpha_i + \epsilon_{i,t} \tag{4}$$

or, equivalently, expressed in terms of month-on-month growth rates:

$$y^*_{i,t} = \gamma_1 \Delta x_{i,t} + \gamma_2 \Delta x_{i,t-1} + \cdots + \gamma_{12} \Delta x_{i,t-11} + \lambda_1 y_{i,t-1} + ... + \lambda_{12} y_{i,t-12} + \alpha_i + \epsilon_{i,t} \tag{5}$$

where

$$\gamma_j = \sum\nolimits_{k=j}^{12} \beta_k,$$

and the model nests the special, and testable, case that the ITS data relate, as the CBI ask, to growth over the last three (previously four) months.

In (4) a firm's response to the ITS depends on both its contemporaneous and lagged turnover growth, as measured by the MPI, and, to account for inertia, its previous qualitative replies to the ITS. We confine attention to lags no greater than twelve months, an assumption supported by our empirical analysis (see below), where we fail to reject the statistical insignificance of lags greater than twelve. α_i is a firm-specific and time-invariant random component such that $\alpha_i \sim N(0, \sigma^2_\alpha)$, which accommodates heterogeneity (across firms) in the thresholds μ_1 and μ_2. $\epsilon_{i,t}$ is a time and firm-specific error term which is assumed to be normally distributed and uncorrelated across firms and uncorrelated with α_i. The variance of $\epsilon_{i,t}$ is set to unity for identification, so that $E(\epsilon_{i,t}, \epsilon_{i,s}) = \rho = \sigma^2_\alpha/(\sigma^2_\alpha + 1)$ for $t \neq s$. Temporal dependence in $y^*_{i,t}$ is captured via the lagged dummy variables $y_{i,t}$ (one for $j = 0$ and one for $j = 2$) and more general dynamics are picked up by considering lags of the explanatory variables. We focus on equation (5), rather than (4), since interpretation is perhaps easier, as it breaks down the cumulative effect of turnover growth over the last k month into the month-on-month impact in the last k months and thereby helps us track down the source of the signal.

To estimate (5), given (3) and the distributional assumption about ϵ_{it}, we derive the probabilities from the conditional distribution of y^*_{it} on Ω_{it}, where Ω_{it} is the information available to firm i up to time t:

$$\begin{aligned} P(y_{it} &= j|\Omega_{it}) = P(\mu_j < y^*_{it} \leq \mu_{j+1}|\Omega_{it}) = \\ &= \Phi\left(\mu_{j+1} - \gamma_1 \Delta x_{i,t} - \cdots - \gamma_{12}\Delta x_{i,t-11} - \lambda_1 y_{i,t-1} + ... - \lambda_{12} y_{i,t-12} - \alpha_i\right) \\ & \quad -\Phi\left(\mu_{j} - \gamma_1 \Delta x_{i,t} - \cdots - \gamma_{12}\Delta x_{i,t-11} - \lambda_1 y_{i,t-1} + ... - \lambda_{12} y_{i,t-12} - \alpha_i\right), \end{aligned} \tag{6}$$

where $\Phi(.)$ denotes the standard normal cumulative density function. To implement the random effects estimator the individual effect α_i can be integrated out and the model estimated by maximum likelihood using Stata 9.0.

4.1 Hypothesis testing: signal and noise

For the qualitative survey data to be clearly "useful", or contain a signal about the quantitative data, the MPI turnover growth terms in (5) have to be statistically significant. Conversely, when there is no informational content to the ITS responses, and they constitute only noise, we should expect the following hypothesis to hold:

$$H_0^1 : \Delta x_{i,t} = \Delta x_{i,t-1} = \cdots = \Delta x_{i,t-11} = 0. \tag{7}$$

Moreover, when a firm does, as instructed by the CBI, base its qualitative reply on its trend growth over the last three (previously four) months, we should expect lags in (5), beyond the horizon of interest (namely three/four months), to be statistically insignificant:

$$H_0^2 : \Delta x_{i,t-k} = 0 \text{ and } y_{i,t-k} = 0, \text{ for all } k > 4. \tag{8}$$

Rejection of H_0^2 implies that firms do not follow the instructions of the CBI, and look further into the past than asked when reporting their trend output growth. Even when firms follow the CBI's instructions we should expect the lagged qualitative responses $y_{i,t-k}$ to be statistically insignificant, but only at lags beyond the horizon of interest. At shorter lags, say $y_{i,t-1}$, we should expect dependence even when firms follows the CBI's advice since y^*_{it} and $y_{i,t-1}$ overlap given that they refer, respectively, to growth over the last four months and growth two month's ago relative to five months ago.

When H_0^2 is rejected we isolate the cause by breaking (8) down into two constituent tests:

$$H_0^3 : \Delta x_{i,t-k} = 0, \text{ for all } k > 4 \tag{9}$$

and

$$H_0^4 : y_{i,t-k} = 0, \text{ for all } k > 4, \tag{10}$$

where H_0^3 tests whether firms base their qualitative response on quantitative information too far back into the past and H_0^4 tests if there is too much (relative to the CBI's question) inertia in firms' qualitative responses. Due to the change in the reference period of the ITS question over our sampling period, we also consider variants of these tests for $k > 3$. Denote the tests for $k > 3$, H_0^{2a}, H_0^{3a} and H_0^{4a}, and those for $k > 4$, H_0^{2b}, H_0^{3b} and H_0^{4b}.

4.2 A quantitative indicator of firm-level output growth constructed from the ITS

To assess whether any signal in the ITS data has *value* to economists we examine the ability of the qualitative data to predict (or nowcast) the firm-level quantitative data. This involves inverting the probit models, via Bayes' Theorem (see also J. Mitchell & Weale (2006)), and constructing an early indicator of the quantitative data based on the qualitative data. Particular interest rests on how useful the ITS data published at time t are at predicting (or, more accurately, nowcasting) the MPI data at time t, Δx_{it}. But given the possibility that the ITS data tell us not just about Δx_{it}, but lags of Δx_{it}, we construct indicators of Δx_{it-k} $(k = 0, 1, 2, ...)$.

Let j_{it}, $(j_{it} = 0, 1, 2)$, denote the qualitative survey response of firm i at time t. Let $f(\Delta x_{it}, \cdots, \Delta x_{it-k} | \{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1})$ denote the prior conditional density for the quantitative data, constructed without reference to the qualitative data. Given the likely correlation of $\Delta x_{it}, \cdots, \Delta x_{it-k}$, this multivariate conditional density is assumed to follow a multivariate normal distribution

$$f(\Delta x_{it}, \cdots, \Delta x_{it-k} | \{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1}) \sim N(\mu, \Sigma). \tag{11}$$

We need to work out the density function of Δx_{it-k} conditional on the firms' observed qualitative survey responses at time t and lags, and conditional on lagged quantitative information $\{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1}$. We denote this density function $f(\Delta x_{it-k} | \{j_{i\nu}\}_{\nu=1}^{t}, \{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1})$. Our indicator, D_{it-k} $(k = 0, 1, 2, ...)$, is then given as:

$$D_{it-k} = E(\Delta x_{it-k} | \{j_{i\nu}\}_{\nu=1}^{t}, \{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1}) \tag{12}$$

$$D_{it-k} = \int_{-\infty}^{\infty} \Delta x_{it-k} f(\Delta x_{it-k} | \{j_{i\nu}\}_{\nu=1}^{t}, \{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1}) d\Delta x_{it-k}. \tag{13}$$

This is the expectation of firm-level output growth at time $t-k$, conditional on the qualitative data up to and including time t but quantitative data only up to and including time $t-k-1$, reflecting the lagged availability of the MPI. The *value* of the qualitative survey data rests on comparison of D_{it-k} against the autoregressive benchmark indicator $E(\Delta x_{it-k}|\{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1})$.

Bayes' theorem states that

$$f(\Delta x_{it-k}|\{j_{i\nu}\}_{\nu=1}^{t},\{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1})=\frac{P(j_{it},\Delta x_{it-k}|\{j_{i\nu}\}_{\nu=1}^{t-1},\{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1})}{P(j_{it}|\{j_{i\nu}\}_{\nu=1}^{t-1},\{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1})} \tag{14}$$

where

$$P(j_{it},\Delta x_{it-k}|\{j_{i\nu}\}_{\nu=1}^{t-1},\{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1})= \tag{15}$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}bf(\Delta x_{it},\cdots,\Delta x_{it-k}|\{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1})d\Delta x_{it}\cdots d\Delta x_{it-k+1},$$

and b integrates out the random effect α_i

$$b=\left(\int_{-\infty}^{\infty}P(j_{it}|\{j_{i\nu}\}_{\nu=1}^{t-1},\{\Delta x_{i\tau}\}_{\tau=1}^{t},\alpha_i)f(\alpha_i)d\alpha_i\right). \tag{16}$$

The denominator of (14) involves integrating Δx_{it-k} out from (15). Note that when $k=0$, since future values of the quantitative data do not enter the probit models, (15) reduces to

$$P(j_{it},\Delta x_{it}|\{j_{i\nu}\}_{\nu=1}^{t-1},\{\Delta x_{i\tau}\}_{\tau=1}^{t-1})=\int_{-\infty}^{\infty}bf(\Delta x_{it}|\{\Delta x_{i\tau}\}_{\tau=1}^{t-1})d\Delta x_{it}. \tag{17}$$

Given $f(\Delta x_{it-k}|\{j_{i\nu}\}_{\nu=1}^{t},\{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1})$, with μ and Σ estimated by least squares, all of the above integrals may be calculated by numerical evaluation. Estimators $\widehat{P}(j_{it}|\{j_{i\nu}\}_{\nu=1}^{t-1},\{\Delta x_{i\tau}\}_{\tau=1}^{t},\alpha_i)$ for $P(j_{it}|\{j_{i\nu}\}_{\nu=1}^{t-1},\{\Delta x_{i\tau}\}_{\tau=1}^{t},\alpha_i)$ are given by substitution of the estimators $\widehat{\sigma}_{\alpha}^{2}$, $\hat{\mu}_j$ ($j=0,...,3$), $\widehat{\gamma}_l$ ($l=0,...,11$) and $\widehat{\lambda}_m$ ($m=1,...,12$) , in (6). Hence, a feasible Bayes estimator $D_{it-k}=\widehat{E}(\Delta x_{it-k}|\{j_{i\nu}\}_{\nu=1}^{t},\{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1})$ may be obtained from (13) by numerical evaluation.

5 Empirical Results

5.1 Signal or noise?

Table 3 summarises the hypothesis tests based on estimation of the dynamic ordered probit model, alongside results from an analogous (auto-regressive distributed-lag, ARDL) model, again with twelve lags, estimated using monthly macroeconomic data (manufacturing growth data from the ONS and the balance statistic from the ITS), from 1985m1 to 2007m12.[5] Consistent with our view that the qualitative survey data provide noisy estimates of the official data (i.e. the measurement error is in the qualitative survey data rather than the official data) and that the official data drive firms' qualitative responses, the ARDL model is estimated with the qualitative data (the balance statistic) as the regressand.

The p-values in Table 3 indicate that the hypothesis of noise, H_0^1, is clearly rejected at traditional significance levels (90%, 95% or 99%) with a p-value less than 1%, for both the firm-level and macroeconomic data. Firms also appear to follow the CBI's instructions and base their qualitative responses on their output movements over the last 3/4 months - this is reflected in our failure to reject H_0^{2a} or H_0^{2b} at traditional significance levels (with p-values of 19% and 23%). Comparison of the firm-level and macroeconomic results reveals that there is also a signal in the ITS data at the aggregate-level, since H_0^1 is again rejected, with a p-value of less than 1%. However, the macroeconomic signal lags the firm-level one.

To provide further indication of the relationship between the ITS and MPI data in the underlying probit model Figure 5 plots the t-statistics of the estimated coefficients on $\Delta x_{i,t}$ to $\Delta x_{i,t-11}$, used as the basis for the hypothesis tests in Table 3, while Figure 6 indicates the relative importance of these estimated coefficients by plotting $\widehat{\gamma}_i / \sum \widehat{\gamma}_i$. Inspection of the lagged dummy variables in the probit models also reveals considerable inertia in firms' qualitative responses. The significance of the lagged dummies, although not reported to save space, declines as k increases. In part, as discussed, this is due to the overlapping nature of their responses rather than genuine state dependence.

The top panel of Figures 5 and 6 again indicates that there is a signal in the ITS data at the firm-level: the ITS data are plainly related to the responses the same firms give to the MPI. Consistent with the hypothesis tests presented in Table 4, the t-values are

[5] Seasonal, time, sectoral and size dummies were included in the probit models but found to be statistically insignificant. This is consistent with the evidence presented in Figures 2 and 3. Further evidence on the reliability of the firm-level results is provided in appendix A.3.

Table 3: Signal or noise? p-values for the hypothesis tests on the firm-level data and the macroeconomic data

Likelihood ratio tests:	Micro	Macro
H_0^1	0.0019	0.0012
H_0^{2a}	0.1867	0.0003
H_0^{2b}	0.2287	0.0025
H_0^{3a}	0.1048	0.0038
H_0^{3b}	0.1679	0.0524
H_0^{4a}	0.3452	0.0185
H_0^{4b}	0.3176	0.0134
$\Delta x_{i,t-k} = 0$, for $k = 12, 13, 14, 15$	0.1731	0.5673

often greater than 1.96, the 95% critical value. As with the raw correlation coefficients seen in Figure 1, Figures 5 and 6 also shows that the signal from the firm-level qualitative survey is strongest (about growth) 2-3 and 4-5 months ago, not contemporaneously. But Figures 5 and 6 do show that the macroeconomic signal lags the firm-level one. The peak t-statistic (in Figure 5) for the macroeconomic data is at $k = 4$, rather than $k = 2$, and, unlike with the firm-level data, one can reject in Table 3, again at traditional significance levels (with a p-value of 0.0038), the null hypothesis that $H_0^{3a} : \Delta x_{t-k} = 0$, for all $k > 3$. This implies that the balance statistic tells us about growth in the economy further back into the past than just the last three months. However, as with the firm-level data, this signal does appear to weaken thereafter, since one can only reject the null hypothesis that $\Delta x_{t-k} = 0$, for all $k > 4$, with a p-value of 0.0524.

5.2 The ITS as an early indicator of the MPI

We assess the predictive power of the indicator D_{it-k} based on the in-sample fit of the probit models. An out-of-sample analysis, which involves splitting the sample, T, in two and estimating the models recursively, is not sensible in this application given that $T = 60$. Moreover, in-sample tests of predictability have been found to have greater power than out-of-sample tests; see Inoue & Kilian (2004).

Before computing D_{it-k} we use the Bayesian Information Criterion (BIC) to reduce the model in equation (5), with up to 12 lags of $y_{i,t}$ and $\Delta x_{i,t}$, to a more parsimonious, but good fitting, model. Table 4 reports the estimation results for this preferred model. As with the plots in Figures 5 and 6 we find that the quantitative data are statistically

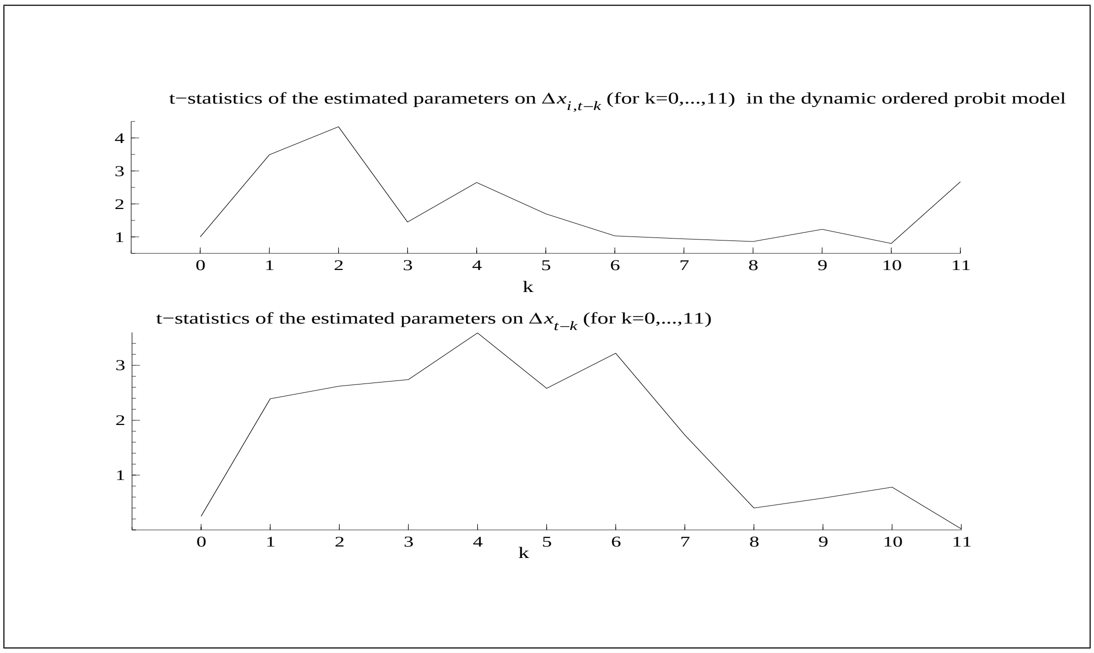

Figure 5: t-statistics of the estimated rolling 1-month MPI turnover growth rates in the firm-level data (top panel) and macroeconomic data (bottom panel)

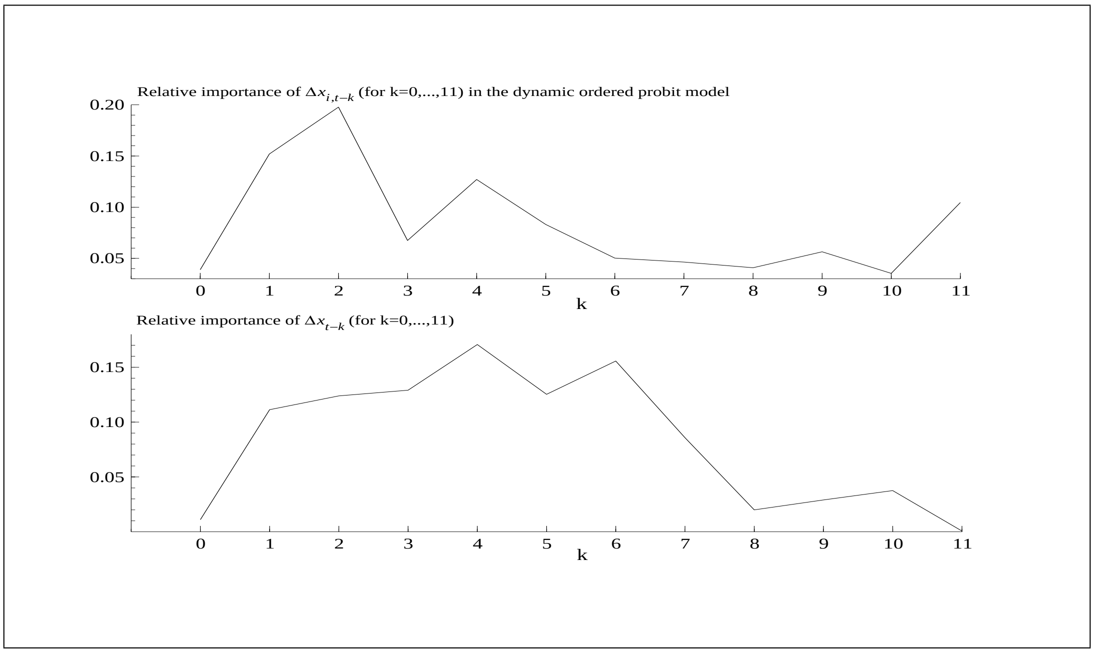

Figure 6: Relative importance of the estimated rolling 1-month MPI turnover growth rates in the firm-level data (top panel) and macroeconomic data (bottom panel)

insignificant at time t and it is lagged values of Δx_{it} that help explain the qualitative data. The lagged qualitative data are again playing a very important role with larger t-statistics than on Δx_{it-1} and Δx_{it-2}. We accept the restriction that $\sigma_{\alpha}^{2} = 0$ and so confine attention to the pooled dynamic ordered probit model.

Table 4: Estimation output for the parsimonious ordered probit model

Explanatory variables	Estimated coeff.	Robust t-stat
$\Delta x_{i,t}$	0.0066	0.17
$\Delta x_{i,t-1}$	0.1294	2.90
$\Delta x_{i,t-2}$	0.1681	3.86
$y_{i,t-1}^{u}$	1.0421	9.67
$y_{i,t-1}^{d}$	-0.8594	-7.69
$y_{i,t-2}^{u}$	0.6064	5.46
$y_{i,t-2}^{d}$	-0.4749	-4.46
lower threshold	-0.8250	-12.71
upper threshold	0.9307	14.41
Number of obs ($\sum_{i=1}^{N} T_i$)		1033
N		96
LR chi2(7)		577.5
$Prob > chi2$		0
Pseudo R^2		0.262
Log likelihood	ll(null)	-1101
	ll(model)	-812
BIC		1687

Table 5 summarises the performance of D_{it-k} by reporting the correlation and Root Mean Squared Error (RMSE) of the indicator, pooled across firms and time, against $\Delta x_{i,t-k}$. Performance of D_{it-k} is distinguished according to the prior density, $f(\Delta x_{it}, \cdots, \Delta x_{it-k} | \{\Delta x_{i\tau}\}_{\tau=1}^{t-k-1})$, chosen to model the MPI data; for robustness, we consider two choices, an AR for each element of $(\Delta x_{it}, \cdots, \Delta x_{it-k})'$ with $p = k+1, ..., 11$ lags and an AR with just a single lag. Parsimonious models are often found to forecast well; e.g., see Clements & Hendry (1998). The final two columns of Table 5 summarise the performance of fitted values from these two models. Since D_{it} is the mean of the posterior density, comparison against the prior mean tells us about the value of the qualitative survey data.

While the RMSE of the indicator D_{it-k} is, in general, lower than the benchmark, purely autoregressive, indicator the differences in Table 5 are extremely small. In addition, the improvement in correlation, in all cases but one, is also very minimal. This indicates that while, as seen in Table 3, the ITS data do contain a signal about the MPI data this signal does not translate into improved predictive power for the MPI relative to benchmark autoregressive forecasts. Consistent with Figure 5, where it was seen that Δx_{it} does not help explain firms' contemporaneous qualitative responses, Table 5 shows that what gains there are to using the posterior indicator, which as indicated are minimal, are confined to lags of Δx_{it}. Therefore, in the case of most interest, we find that conditioning our forecast of the latest MPI data, Δx_{it}, on the latest ITS data, which are available ahead of the MPI data, does not deliver more accurate nowcasts.

Table 5: Predictive performance of the ITS indicator of firm-level MPI turnover growth

	D_{it}: Posterior Mean		Prior Mean	
	AR(11)	AR(1)	AR(11)	AR(1)
Corr with $\Delta x_{i,t}$	0.582	0.394	0.582	0.394
RMSE	0.807	0.912	0.807	0.912
	D_{it-1}: Posterior Mean		Prior Mean	
	AR(10)	AR(1)	AR(10)	AR(1)
Corr with $\Delta x_{i,t-1}$	0.528	0.396	0.519	0.386
RMSE	0.854	0.923	0.859	0.928
	D_{it-2}: Posterior Mean		Prior Mean	
	AR(9)	AR(1)	AR(9)	AR(1)
Corr with $\Delta x_{i,t-2}$	0.511	0.166	0.508	0.381
RMSE	0.859	0.994	0.862	0.997

Table 6 shows that a similar picture again emerges at the macroeconomic level: conditioning autoregressive (in-sample) forecasts of Δx_{t-k} on contemporaneous and lagged values of the balance statistic appears to deliver, at best, minimal gains, with the greater gains again confined to lags of Δx_t. The posterior mean nowcasts in Table 6 are based on estimation of an ARDL$(11-k, q)$ model; this involved, as with the widely used quantification approach of Pesaran (1984, 1987) (e.g. see Driver & Urga (2004)), regressing the official data (manufacturing output growth Δx_t) on $p = k+1, ..., 11$ lags (Δx_{it-p}) and current and $(q-1)$ lagged values of the balance statistic from the ITS. We note that the alternative, considered above, of running the regression the other way round, and then inverting it to obtain posterior mean nowcasts of Δx_{it-k}, delivered even worse performing posterior mean nowcasts.

No doubt these macroeconomic results are sensitive to the sample period chosen for analysis. Our point is simply that given the firm-level results in Table 5 it should not perhaps be a surprise to find that it is hard at the macroeconomic level, systematically over time, to beat autoregressive (benchmark) forecasts.

Table 6: Predictive performance of the ITS balance statistic for aggregate manufacturing output growth relative to a benchmark (prior) AR model in manufacturing output growth

	Posterior Mean of Δx_t		Prior Mean	
	ARDL(11,3)	ARDL(1,3)	AR(11)	AR(1)
Corr with Δx_t	0.403	0.345	0.397	0.302
RMSE/sd(Δx_t)	0.913	0.937	0.916	0.951
	Posterior Mean of Δx_{t-1}		Prior Mean	
	ARDL(10,3)	ARDL(1,3)	AR(10)	AR(1)
Corr with Δx_{t-1}	0.427	0.371	0.399	0.305
RMSE/sd(Δx_{t-1})	0.902	0.927	0.915	0.951
	Posterior Mean of Δx_{t-2}		Prior Mean	
	ARDL(9,3)	ARDL(1,3)	AR(9)	AR(1)
Corr with Δx_{t-2}	0.435	0.383	0.399	0.308
RMSE/sd(Δx_{t-2})	0.899	0.922	0.915	0.949

6 Conclusion

This paper first tests the reliability of qualitative business survey data against official quantitative data at the firm level, as well as, which is more common, at the macroeconomic level. The firm level exercise involved construction of a unique dataset, which involved matching a panel of firms' responses to the qualitative and quantitative surveys. This new dataset is then analysed to provide a definitive means of assessing the informational content of qualitative business surveys of the type routinely analysed by economists. These types of surveys have the advantage of timeliness relative to official surveys. Moreover, they ask firms a range of questions not posed in official surveys. But, as this paper explains, for those questions, like the retrospective output question, which have a natural counterpart in official surveys, it is possible to test the informational content of the qualitative data against the quantitative data at the firm-level. Clearly the approach we suggest could, with the cooperation of data producers, be applied to related surveys in other countries.

Our application to firm-level data from the ITS finds that the retrospective qualitative data are plainly related to the responses the same firms gave to the MPI. Firms also appear to follow the CBI's instructions quite closely. However, having introduced a novel means of inferring the official quantitative data from the qualitative data, we find that conditioning autoregressive forecasts of the MPI data on contemporaneous values of the ITS data does not improve inference. Nevertheless, the statistical links we do observe do raise confidence in the information content of those data which have no counterpart in official surveys.

7 Appendix

7.1 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. The response rate to the MPI is quite high, averaging around 82%. This corresponds to around 8000 responses each month. In the case of non-response, imputation via a "near neighbour" approach is adopted by the ONS.

7.2 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. The same sample of firms are questioned each month and firms are only removed from the sample if they go out of business or when they opt not to respond, which may be because they no longer maintain their CBI membership (although membership of the CBI is not a prerequisite for participation in the ITS since the CBI do not demand removal from the ITS if firms leave the CBI) or if they choose to have a response holiday. There is a gradual increase in the sample size as new participants are added to the survey in the areas where old participants resigned. 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. There were over our sample period, see Table 1, around 800 respondents to the ITS survey each month.

7.3 Reliability of empirical results

Only those firms with at least twelve consecutive observations (or thirteen in the level of turnover) are included in the model, (5), as estimated in Table 3 whose t-values are shown in Figure 5. This means our sample drops from the 807 firms seen in Table 1 to 96 firms, with these 96 firms having 1033 observations between them (i.e., pooled across time). To provide some check on how reliable results from this sub-panel are, first in Figure 6 we simply compare the polyserial correlations between the full and sub-panels. When the 96 firms are a random sample from the 807 we should expect statistics derived from both not to differ significantly. Therefore, it is reassuring that the polyserial correlations from the sub-panel falls within the 95% confidence interval around the correlation estimates from the full panel.

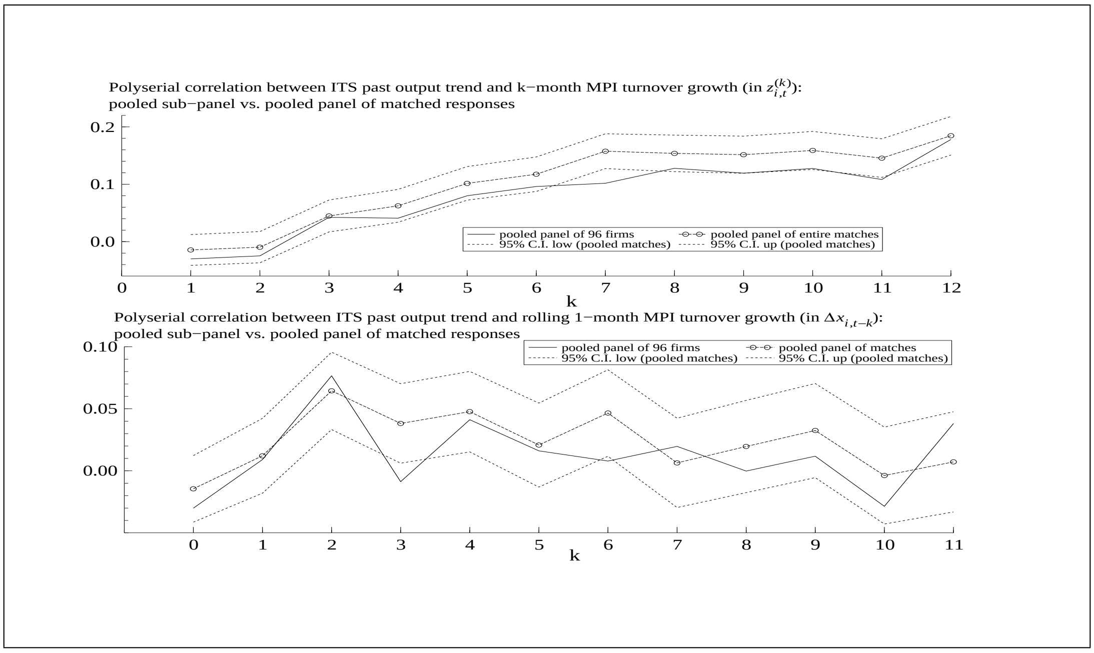

Figure 7: Pooled polyserial correlation between the firm-level qualitative and quantitative data using the full panel and the sub-panel of 96 firms

To provide a further indication of how representative results from the sub-panel are, we consider the variable addition tests of Verbeek & Nijman (1992). These involve adding a test variable to the random effects ordered probit model considered in Table 3. This test variable is a count variable characterising the number of times a firm is present in the matched dataset (summarised in Table 1) across the 60 months. When the sub-panel is a random sample from the full panel we should not expect firms' qualitative responses to be associated with the number of times a firm is present in the matched dataset. We find a p-value of 0.979 in the dynamic ordered probit model considered in Table 3, which supports the view that the results from the 96 firms are representative of those from the full panel. In addition, we re-estimated the parsimonious ordered probit model in Table 4 but, since we no longer test this model against more general alternatives which reduced the number of firms available to 96, let an additional 285 firms enter the sample. A Hausman test could not reject the null of equivalence between the two estimators at 95% [the p-value was 0.155], lending further support to the view that the results in Tables 3 and 4 are free from sample selection.

References

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

Clements, M. P. & Hendry, D. F. (1998), *Forecasting Economic Time Series*, Cambridge University Press: Cambridge.

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

Inoue, A. & Kilian, L. (2004), 'In-sample or out-of-sample tests of predictability: Which one should we use?', *Econometric Reviews* **23**, 371–402.

J. Mitchell, R. S. & Weale, M. (2006), 'A bayesian indicator of manufacturing output from qualitative business panel survey data'. National Institute of Economic and Social Research Discussion Paper No.261 (Revised)).

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

Olsson, U., Drasgow, F. & Dorans, N. J. (1982), 'The polyserial correlation coefficient', *Psychometrika* **47**, 337–347.

Pesaran, M. H. (1984), Expectations formation and macroeconomic modelling, *in* P. Malgrange & P. Muet, eds, 'Contemporary Macroeconomic Modelling', Blackwell, Oxford, pp. 27–55.

Pesaran, M. H. (1987), *The limits to rational expectations*, Basil Blackwel, Oxford.

Pesaran, M. & 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.

Verbeek, M. & Nijman, T. (1992), 'Testing for selectivity bias in panel data models', *International Economic Review* **33**, 681–703.