ESA/STAT/AC. 223/S4.3

Third International Seminar on Early Warning and Business Cycle Indicators

17 – 19 November 2010
Moscow, Russian Federation

Qualitative Business Surveys: Signal or Noise?

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

Qualitative Business Surveys: Signal or Noise?

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

June 4, 2010

Abstract

This paper identifies the information content at the firm level of qualitative business survey data by examining the consistency between these data and the quantitative data provided by the same respondents to the UK's Office for National Statistics 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 the firm-level quantitative data.

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

*Address for correspondence: National Institute of Economic and Social Research, 2 Dean Trench Street, Smith Square, London, SW1P 3HE, U.K. Tel: +44 (0) 207 654 1926. Fax: +44 (0) 207 654 1900. E-Mail: s.lui@niesr.ac.uk; j.mitchell@niesr.ac.uk and m.weale@niesr.ac.uk.

1 Introduction

The purpose of this paper is to assess, at the firm level, the relationship between the qualitative business survey data produced by the Confederation of British Industry (CBI) collected in its Industrial Trends Survey (ITS) and related quantitative data collected by the Office for National Statistics (ONS) in its Monthly Production Inquiry (MPI) and used for the construction of the Index of Industrial Production (IoP). In the ITS 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 database of business surveys for the European Union (see http://ec.europa.eu/economy_finance/db_indicators/surveys9185_en.htm).

Interest focuses on the CBI survey and similar surveys as output indicators because, although qualitative by nature, they are seen as more timely than ONS data. In recent years this interest has increased as a result of doubts about official output data (see Ashley et al. (2005)). However past studies of the relationship with official output data have largely relied on comparisons between summary statistics from the two sources. From the qualitative survey the proportion of firms reporting a rise in activity less the proportion reporting a decline is computed, normally after weighting the respondents by an indicator of their size. This figure is known as the balance statistic and is compared with the percentage change in output reported by the ONS. Driver & Urga (2004) and Pesaran & Weale (2006) provide a review of such comparisons. At best, there have been comparisons between aggregate ONS data and the panel of firm-level qualitative 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. (2002, 2005, 2010)). Thus, on the basis of these studies it is difficult to say how firmly the perceived relationships between the two data sets are based.

In this paper, by contrast, we compare the individual responses provided to the CBI with those collected by the ONS on a firm-by-firm basis, test whether a relationship exists and, if so, identify its form. Then, 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 the firm-level quantitative data. Our focus is on the manufacturing sector since this is what the ITS covers, but, with suitable data a similar analysis could be carried out for other parts of the economy.

The plan of the remainder of the paper is as follows. Section 2 describes the matched panel dataset, discussing first the matching process and then examining the statistical properties of the matched dataset. Section 3 provides some simple descriptive evidence summarising the relationship between firms' qualitative and quantitative responses. Section 4 then supplements this with more formal econometric modelling which lets us test whether the qualitative survey data contain a signal about the 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

2.1 Background

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 ITS is a voluntary survey open to both CBI members and non-member companies. The ITS asks firms many questions, only some of which, like those on output, are 'verifiable', that is testable against official data. Other questions in the ITS, such as "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 verified against official ONS data.

In this paper we focus on the question from the ITS 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" (and "not applicable"). Firms are also asked to indicate how many employees they have within four size-bands: 0-199; 200-499; 500-4999 and 5000+ employees and we use this to provide extra information on our data. From the start of the survey in 1958 until June 2003 firms were asked about the trend in output over the last four months. From July 2003 onwards this was changed to three months as indicated above.

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 matched the CBI data set to its own MPI for the five years 2000-2004, inclusive. The MPI is a sample-based survey which asks close to 9000 firms each month for quantitative information on their turnover values in the month (see ONS (2005) for further details). Firms are also asked to indicate how many employees they have. The MPI uses stratified random sampling, stratifying the population by industry and employment. 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 within 18 working days into the following month. In the event of non-response by firms which have responded previously, the ONS imputes a current value based on the average movement of other firms within the same industry, using mean imputation rather than nearest neighbour. Over the five years of our sample period, pooled across firms about 11% of responses to the turnover question in the MPI sample are imputed.

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. The ONS measures this by means of a separate Quarterly Inventories Inquiry and uses a small monthly inquiry to supplement these data so as to produce the monthly figures needed to deliver the IoP (ONS 2007). The adjustments are applied at an industry level and not at a firm level, however. 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. As we explain in Section 4, we should still expect the MPI and ITS responses to be related despite the effects of stock changes.

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.

The IoP is 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 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 as a timely source of information about the current state of the economy.

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 desire to examine this question motivates the empirical work below that tests the informational content of these ITS firm-level data.

2.2 The matching process

The matched ITS/MPI dataset, the focus of our statistical and econometric analysis, was constructed 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 (ONS 2009). It covers all parts of the economy, but excludes some of the very small businesses. They are usually those of the self-employed, those without employees and with low sample turnover. Some non-profit making organisations are also excluded. Samples for the MPI are derived from the IDBR via a Permanent Random Number system (Ohlsson 1995), which allows gradual rotation of the sample within each stratum for each four-digit industry. However, the sampling fraction for the 'largest' firms, typically with more than 150 employees, is 100% and the firms in this stratum stay in the sample permanently. In Table 1 we provide statistics on the number of responding firms in each year in the MPI, the ITS, the ITS firms matched to the IDBR and the subset of these which matches the MPI data set. These are provided for the start of each year and also at the end of our data set.

Column A summarises the responses to the MPI at the start of each year and also shows the total number of distinct respondents over the five-year period. Column B in Table 1 provides similar data for the ITS. As well as ignoring the five firms who opted out, and whose response pattern we do not know, these figures also omit responses to twenty-five ID numbers in the ITS data set used for anonymous responses; these IDs may represent responses from more than one firm.

Text matching based on common variables, specifically the names, addresses and postcodes of firms, was used to match these CBI firms against those firms on the ONS's IDBR. Of the 2584 different firms who replied at least once to the ITS 2120 (82%) were initially matched against the IDBR; we are, however, not able to allocate these across individual months. Of these 2120 firms, and defining a 'definite' match as when the ONS is at least 80% confident in the match, column C indicates there was such a match between the ITS firms and the IDBR for 1895 firms, i.e., for about 90% of firms. For the remaining 464 firms 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

Table 1: The number of firms in the raw datasets and the matched response dataset

Year	A MPI firms	B ITS Firms	C ITS Firms uniquely Matched to IDBR	D Matched ITS -MPI Firms
Jan 2000	9005	614	445	152
Jan 2001	9046	783	575	163
Jan 2002	8959	841	623	198
Jan 2003	8916	820	608	158
Jan 2004	8913	796	588	176
Dec 2004	8641	750	585	160
Total No. Distinct Respondents	28033	2584	1895	807
Sample Turnover Rate (% p.a.)	44%	48%	47%	77%

the ITS; on the advice of the ONS these firms were therefore dropped from our analysis to minimise the risk of matching errors. The ONS quantify their confidence in a match based on the percentage of these common variables which are matched successfully across the two surveys using matching software produced by Search Software America, with the name and address of a firm, which themselves can be compounds, assigned a higher weight than the postcode.

The subset of ITS firms uniquely matched to the IDBR which are also sampled by the MPI could then be extracted in the second stage of the matching. 807 different firms gave at least one contemporaneous matched response to the ITS and MPI. Column D in Table 1 shows that, on average across years, there are about 170 different firms present in the matched dataset each month. 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% although the average monthly match rate was 22%. There is a total of 10254 responses provided by these 807 firms. About 4% are imputations provided by ONS. We have not distinguished these from the hard data in the dataset.

The sample turnover rates shown in the last row of Table 1, and discussed further below, are calculated as follows. Consider first the case where $Y_{m,1}$ indicates the number of firms in the dataset of interest (column A-D) responding in the first month of year m and we have data for M years ($m = 1, ..., M$). $\tilde{Y}$ is the total number of firms responding in all M years of the survey. We can then define the sample turnover rate as the value of r which satisfies

$$\sum_{m=1}^{M} Y_{m,1} - (1-r) \sum_{m=1}^{M-1} Y_{m,1} = \tilde{Y}. \tag{1}$$

However, in our case our last observation is for the last month of year $M-1$ rather than the first month of year M; we denote this $Y_{M-1,12}$. We therefore define the sample turnover rate, r, as that value which satisfies

$$\sum_{m=1}^{M-1} Y_{m,1} + Y_{M-1,12} - (1-r) \sum_{m=1}^{M-2} Y_{m,1} - (1-r)^{\frac{11}{12}} Y_{M-1,1} = \tilde{Y} \tag{2}$$

reflecting the fact that the interval between $Y_{M-1,1}$ and $Y_{M-1,12}$ is only eleven months.

2.3 Statistical properties of the matched ITS-MPI dataset

We summarise the characteristics of the matched dataset with a number of statistics. First we explore how many respondents in the matched data set provide responses in more than one period; since we are interested in an econometric analysis involving lag terms we are particularly interested in knowing how many respondents provide runs of consecutive responses. And secondly we discuss the extent to which the matched sample is representative of the economy as a whole.

The total number of times a given firm, firm i ($i = 1, ..., N$), 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 the 807 firms have at most 3 matched (contemporaneous) responses; 50% have at most 8 matched responses and 75% have at most 16 matched responses.

However, consecutive responses are much less frequent. In the matched dataset 96 firms reply to at least twelve consecutive observations (or thirteen in the level of turnover). These 96 firms have 1033 observations between them (i.e., pooled across time).

A feature of Table 1 is that the sample turnover rate in the matched sample is much higher than in either the ITS or MPI. However, if we focus on the retention rate, as one minus the sample turnover rate, we note that the retention rate for the matched data set is, at 23%, not much lower than the product of the retention rates of the MPI and the ITS, which is 29%. The latter is the retention rate which would be expected from large samples if sample turnover were driven by independent processes in the two surveys. The lower actual

Table 2: Proportion of firms in different employment size bands (averaged across years 2000-2004 inclusive)

Band	1	2	3	4
Employment Size	0 to 199	200 to 499	500 to 4999	$\geq$ 5000
MPI	75.40%	16.6%	7.8%	0.2%
ITS	76.50%	14.8%	7.8%	0.9%
Matched Firms	53.57%	32.2%	13.6%	0.7%
96-firm Sub-panel of Matched Firms	42.52%	45.0%	12.0%	0.5%

retention rate implies that a firm is more likely to remain in the ITS sample if it drops out of the MPI sample than if it remains in and *vice versa*; we have been unable to identify reasons for this and note that, to the extent that sample turnover is driven by a common cause such as survival discussed below, the retention rate in the matched data set would be expected to be higher than that calculated on the assumption of independence.

Table 2 indicates the average (across the years 2000-2004, inclusive) size of firms in the matched ITS-MPI dataset, including the sub-panel of 96 firms with at least 12 consecutive responses. Table 2 also reports the average size of firms in the ITS and MPI. Since the size of a firm can change over time, in a given year we take an average of the available monthly observations. For the MPI these observations are quantitative and we take the mean, while for the ITS they are categorical and we take the mode. Taking the modal value from the MPI led to similar results.

Table 2 shows that the MPI and ITS have similar characteristics, although the ITS does sample more large firms. But the matched ITS-MPI dataset contains a far bigger proportion of large firms than the MPI. Around three-quarters of firms in the MPI have fewer than 200 employees, as opposed to around one half in the matched dataset. The proportion of firms with fewer than 200 employees is even lower in the 96 firm sub-panel.

This over-representation of large firms in the matched dataset is explained by a number of factors. First, the stratified sample design of the MPI ensures that larger firms are more likely to remain in the sample and thus contribute to both the matched dataset and the subset of matched firms with at least 12 consecutive monthly responses. Secondly, it is easier to match the larger (well-known) firms than the smaller firms. Recall the ONS classify a match as "definite" when their matching algorithms have 80% confidence in the match. For those firms that make the matched dataset, 13% of firms with fewer than 200 employees have a confidence score less than 95%; whereas only 8% of firms with more than 500 employees have a score less than 95%.

To shed further light on why the bulk of firms (807-96=711 firms) do not survive the twelve-month period, leading to the high sample turnover rate for the matched dataset seen in Table 1, we looked at the ONS's Business Structure Database (BSD). The BSD contains successive (yearly) vintages of the IDBR, and provides an historical record of the life-span and structure of firms. We found that across the five years 14%-21% of these 711 firms dropped out of the matched dataset in the year following their presence in the matched dataset because of death or demographic change, such as a takeover, merger or re-structuring including a change in ownership. This implies that the high sample turnover rate observed for the matched dataset is explained, in large part, not by death or demographic change, but living firms not replying to the MPI and ITS simultaneously.

In summary, this evidence suggests that the matched ITS-MPI dataset, with its strong bias towards large firms, is not a random sample from the population of UK manufacturing firms, represented by the MPI. But this does not mean that the matched dataset picked up firms that were either particularly *good* or *bad* at replying to the ITS. 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. It is therefore important in the econometric analysis below to undertake econometric tests for sample selection; an issue which we explore in Section 5.1.4. Beyond this, 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.

2.4 Further statistical issues

Before comparison of the two surveys can be carried out using these matched data a range of statistical issues needs to be addressed. The firm-level MPI turnover data are measured in current prices and have to be converted to constant prices by means of appropriate output price indices. The ONS classifies each MPI respondent to a particular industry at a 4-digit level in the Standard Industrial Classification which is the most detailed level of the hierarchical classification. This classification is based on the nature of the firm's principal product and the ONS has provided an output price index for each 4-digit SIC category with the price indices selected on the basis of each firm's current industrial classification. We have used these price indices to convert the turnover data to measures at constant prices.

The firm-level 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 level, we prefer to address seasonality by using appropriate dummy variables in the models we use to assess the relationship between the MPI and ITS data. An exception is the case where we consider differences over 12 months since in this case the seasonal effects are removed. Specifically, let $x_{i,t}$ denote the volume of turnover for firm i for time (month) t ($t = 1, ..., T$, where $T = 60$ months in our application), computed from the MPI. The k-month growth rate is defined as

$$z_{i,t}^{(k)} = \ln x_{i,t} - \ln x_{i,t-k}, \ (k = 1, ..., 12), \tag{3}$$

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

$$\Delta x_{i,t-k} = \ln x_{i,t-k} - \ln x_{i,t-k-1} = z_{i,t-k}^{(1)}, \ (k = 0, ..., 11). \tag{4}$$

In addition, 2.5% Winsorisation on each of the upper and lower tails of the distribution of the turnover growth rates, pooled across firms for each period, is carried out prior to analysis to mitigate the possible effects of outliers. 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.

3 Descriptive statistics

Table 3 shows, pooled across firms and time with $\sum_{i=1}^{N} T_i$ denoting the size of the pooled sample, the average (quantitative) turnover growth rate in both the Winsorised and non-Winsorised MPI datasets, conditional on firms' qualitative answers to the ITS. Results are shown both for the full matched panel set and the sub-sample of 96 firms from the full sample with 12 consecutive responses which we examine further below. In both cases, we see that when relating the ITS to turnover growth over the last month, i.e. $\Delta x_{i,t}$, the firms who replied to the ITS saying their turnover had gone up did not experience turnover growth greater than that of other firms. For example, over the full matched non-Winsorised sample, firms reporting falls, in reality, saw their mean deflated turnover rise by 0.53% while firms reporting rises saw their mean turnover fall, but only slightly, by -0.08%.

However, when relating firms' qualitative answers to $\Delta x_{i,t-1}$ we do see that firms reporting rises experienced greater mean growth than firms reporting falls, although only for $\Delta x_{i,t-2}$ do we also see those firms who reported their output had not changed experienced turnover growth in between the means for those reporting rises and those reporting falls. In addition,

Table 3: Mean growth rate (×100) for MPI turnover given firms' qualitative responses to the ITS

Prior to Winsorisation (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
SD of Mean	0.86	0.67	0.76	0.97	0.79	0.86	1.00	0.79	0.90
$\sum_{i=1}^{N} T_i$	1977	2948	1862	1464	2246	1389	1388	2160	1352
$2\frac{1}{2}$% upper and lower tails Winsorised (full matched panel)									
	Δx_{it}			Δx_{it-1}			Δx_{it-2}		
	down	same	up	down	same	up	down	same	up
Mean	0.65	0.46	0.06	-0.15	0.72	0.79	-1.20	0.13	2.64
SD of Mean	0.56	0.44	0.56	0.64	0.49	0.63	0.68	0.51	0.65
$\sum_{i=1}^{N} T_i$	1977	2948	1862	1464	2246	1389	1388	2160	1352
Prior to Winsorisation (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
SD of Mean	3.87	2.39	2.51	3.66	2.49	2.51	3.63	2.42	2.52
$\sum_{i=1}^{N} T_i$	264	469	300	264	469	300	264	469	300
$2\frac{1}{2}$% upper and lower tails Winsorised (sub-panel of 96 firms)									
	Δx_{it}			Δx_{it-1}			Δx_{it-2}		
	down	same	up	down	same	up	down	same	up
Mean	1.23	0.22	-0.41	-1.88	1.25	0.38	-2.04	-1.15	3.14
SD of Mean	1.62	1.11	1.44	1.60	1.16	1.46	1.55	1.15	1.44
$\sum_{i=1}^{N} T_i$	264	469	300	264	469	300	264	469	300

Notes: SD of Mean denotes the standard deviation of the sample mean. $\sum_{i=1}^{N} T_i$ denotes the size of the sample pooled across firms and time. The 96 firm sub-panel consists of those firms in the matched dataset with at least twelve consecutive observations.

at $\Delta x_{i,t-2}$ we observe larger differences between the turnover growth rates of optimistic and pessimistic firms. The only differences between the average growth rates in the different categories which are statistically significant at a 95% level (with t-values>1.96) are those for period $t-2$ between the mean growth rate for those firms reporting down and up, or the same and up, from the full sample, whether Winsorised or not and from the sub-panel of 96 firms after Winsorisation. These calculations are made using the variance of the difference between the means computed as the sums of the variances of the means for the two categories compared. The variance estimates can be added due to independence since the sample means for each categorical response are based on disjoint sets of observations.

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 analysis takes no specific account of the sampling design of either the MPI or the ITS.

The ITS asks each firm, i $(i = 1, \cdots, N_t)$ at time t $(t = 1, \cdots, T)$, 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). Firms can, and a very small number do, also respond 'not applicable'. We ignore these firms below.

We then assume that there is a continuous latent variable $y^*_{i,t}$ that triggers firm i's categorical response at time t via the following observation rule:

$$y_{i,t} = j \text{ if } \mu_j < y^*_{i,t} \leq \mu_{j+1},\ j = 0, 1, 2, \tag{5}$$

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

The latent variable $y^*_{i,t}$, as is set out in our model in equation (6) or (7), is then assumed to depend on both a firm's contemporaneous and lagged turnover growth, as measured by the MPI, and, to account for inertia, its previous qualitative replies to the ITS. Temporal dependence in $y^*_{i,t}$ is captured via the lagged dummy variables $y^{(j)}_{i,t}$ (one for $j = 0$ and one for $j = 2$, with $j = 1$ excluded to avoid collinearity). These take a value 1 if $y_{i,t} = j$ and 0 otherwise. More general dynamics are picked up by considering lags of the explanatory variables. We start with lags of up to twelve months in $z_{i,t}$ and $\left(y^{(0)}_{i,t}, y^{(2)}_{i,t}\right)$ and assume that $y^*_{i,t}$ provides noisy estimates of the quantitative data, $z^{(k)}_{i,t}$; thus we consider the following general model:

$$
\begin{aligned}
y^*_{i,t} &= \beta_1 z^{(1)}_{i,t} + \beta_2 z^{(2)}_{i,t} + \cdots + \beta_{12} z^{(12)}_{i,t} + \lambda^{(0)}_1 y^{(0)}_{i,t-1} + ... + \lambda^{(0)}_{12} y^{(0)}_{i,t-12} \\
&\quad + \lambda^{(2)}_1 y^{(2)}_{i,t-1} + ... + \lambda^{(2)}_{12} y^{(2)}_{i,t-12} + \alpha_i + \epsilon_{i,t}
\end{aligned} \tag{6}
$$

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

$$
\begin{aligned}
y^*_{i,t} &= \gamma_1 \Delta x_{i,t} + \gamma_2 \Delta x_{i,t-1} + \cdots + \gamma_{12} \Delta x_{i,t-11} + \lambda^{(0)}_1 y^{(0)}_{i,t-1} + ... + \lambda^{(0)}_{12} y^{(0)}_{i,t-12} \\
&\quad + \lambda^{(2)}_1 y^{(2)}_{i,t-1} + ... + \lambda^{(2)}_{12} y^{(2)}_{i,t-12} + \alpha_i + \epsilon_{i,t}
\end{aligned} \tag{7}
$$

where $\gamma_j = \sum_{k=j}^{12} \beta_k$, 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.

α_i is a firm-specific and time-invariant random component such that $\alpha_i \sim N(0, \sigma^2_\alpha) = f(\alpha_i)$, 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. We focus on equation (7), rather than (6), 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. Appendix A explains that this general model, (7) can also be motivated as the solution of a two-equation model which accommodates the potential endogeneity of sales/turnover growth, as long as a sufficient number of lagged terms in $\Delta x_{i,t}$ are included.

The assumption that $\epsilon_{i,t}$ is independent across firms, i, while commonly made in applied work, is restrictive. Independence rules out common shocks that affect all firms' responses. In the spirit of Pesaran (2006), who suggested augmenting the panel data model with cross-sectional averages of the dependent and independent variables, we seek to accommodate cross-sectional dependence by letting $\epsilon_{i,t}$ depend on the balance statistic, Bal_t, and aggregate output growth, Δx_t, (and in principle their lags), such that

$$
\epsilon_{i,t} = \delta_1 Bal_t + \delta_2 \Delta x_t + \upsilon_{i,t}, \tag{8}
$$

where $\upsilon_{i,t}$, an idiosyncratic shock, is normally distributed white noise and uncorrelated across firms; Bal_t is defined as the proportion of firms reporting a rise in output less the proportion reporting a fall and Δx_t is aggregate growth. Firms' sentiment as characterised by the ITS, in other words, may have a common collective component, as well as individual components;

see also Lui et al. (2009).

The series for Bal_t is computed from the ITS and available, together with similar data for other European countries, from the European Commission's database referred to above. Firms' responses are weighted by sales data provided to the CBI and the aggregate is seasonally adjusted. The series for Δx_t is the first difference of the logarithm of the manufacturing component of the IoP (series CKYY on the ONS database) and takes account of changes in stocks of finished goods.

To estimate (7), given (5) and the distributional assumption about $\upsilon_{i,t}$, we derive the probabilities from the conditional distribution of $y^*_{i,t}$ on $\Omega_{i,t}$, where $\Omega_{i,t}$ is the information available to firm i up to and including time t:

$$
\begin{aligned}
P(y_{i,t} &= j|\Omega_{i,t}) = P(\mu_j < y^*_{i,t} \leq \mu_{j+1}|\Omega_{i,t}) = P_{ji,t} = \\
&= \Phi\left(\begin{array}{c} \mu_{j+1} - \gamma_1 \Delta x_{i,t} - \cdots - \gamma_{12}\Delta x_{i,t-11} - \lambda_1^{(0)} y_{i,t-1}^{(0)} - \ldots - \lambda_{12}^{(0)} y_{i,t-12}^{(0)} \\ -\lambda_1^{(2)} y_{i,t-1}^{(2)} - \ldots - \lambda_{12}^{(2)} y_{i,t-12}^{(2)} - \delta_1 Bal_t - \delta_2 x_t - \alpha_i \end{array}\right) \\
&\quad -\Phi\left(\begin{array}{c} \mu_{j} - \gamma_1 \Delta x_{i,t} - \cdots - \gamma_{12}\Delta x_{i,t-11} - \lambda_1^{(0)} y_{i,t-1}^{(0)} - \ldots - \lambda_{12}^{(0)} y_{i,t-12}^{(0)} \\ -\lambda_1^{(2)} y_{i,t-1}^{(2)} - \ldots - \lambda_{12}^{(2)} y_{i,t-12}^{(2)} - \delta_1 Bal_t - \delta_2 x_t - \alpha_i \end{array}\right),
\end{aligned} \tag{9}
$$

where $\Phi(.)$ denotes the standard normal cumulative density function. This is a two level probit model, where level 1 are time-series observations within level 2 of firms.

The log-likelihood function, following Butler & Moffit (1982), is then given as

$$
\ln L = \sum\nolimits_{i=1}^{N} \ln \int_{-\infty}^{\infty} \prod\nolimits_{t=1}^{T} \left(P_{0i,t}^{y^0_{i,t}} P_{1i,t}^{y^1_{i,t}} P_{2i,t}^{y^2_{i,t}} \right) f(\alpha_i) d\alpha_i. \tag{10}
$$

where $y^j_{i,t} = 1$ if $\mu_j < y^*_{i,t} \leq \mu_{j+1}$, 0 otherwise, $j = 0, 1, 2$ and N is the total number of different firms present over time ($t = 1, ..., T$). The Stata programme reoprob was used for maximisation.

Under the above assumptions, maximisation of (10) yields consistent estimates ($N, T \to \infty$) for the coefficients denoted: $\widehat{\sigma}^2_\alpha$, $\hat{\mu}_j$ ($j = 0, ..., 3$), $\widehat{\gamma}_l$ ($l = 0, ..., 11$), $\widehat{\lambda}^j_m$ ($m = 1, ..., 12$, $j = 0$ and 2), $\widehat{\delta}_1$ and $\widehat{\delta}_2$.

To validate assumptions implicit in (7) and (8) we carry out diagnostic tests on the generalised residuals from (7) and (8). The generalised residuals of Gourieroux et al. (1987)

are given by

$$E(v_{i,t} \mid y_{i,t}=j, \Delta x_{i,t},..., \Delta x_{i,t-11}, y^{(0)}_{i,t-1},..., y^{(0)}_{i,t-12}, y^{(2)}_{i,t-1},..., y^{(2)}_{i,t-12}, Bal_t, x_t) = \frac{\phi\left(\widehat{\mu}_j - \widehat{\Theta}\right) - \phi\left(\widehat{\mu}_{j+1} - \widehat{\Theta}\right)}{\Phi\left(\widehat{\mu}_{j+1} - \widehat{\Theta}\right) - \Phi\left(\widehat{\mu}_j - \widehat{\Theta}\right)} \quad (11)$$

where ϕ denotes the standard normal density function,

$$\widehat{\Theta} = \widehat{\gamma}_1 \Delta x_{i,t} + \widehat{\gamma}_2 \Delta x_{i,t-1} + \cdots + \widehat{\gamma}_{12} \Delta x_{i,t-11} + \widehat{\lambda}^{(0)}_1 y^{(0)}_{i,t-1} + ... + \widehat{\lambda}^{(0)}_{12} y^{(0)}_{i,t-12} + \widehat{\lambda}^{(2)}_1 y^{(2)}_{i,t-1} + \widehat{\lambda}^{(2)}_{12} y^{(2)}_{i,t-12} + \widehat{\delta}_1 Bal_t + \widehat{\delta}_2 \Delta x_t \quad (12)$$

and a circumflex denotes that unknown parameters are replaced by their maximum likelihood estimates. Diagnostic tests can then be carried out on the generalised residuals; e.g., see Machin & Stewart (1990). We test their normality (using a Jarque-Bera test) and their cross-sectional independence. We use the cross-sectional independence test developed by Hsiao et al. (2007) for use with nonlinear panel data models. The CD test statistic is based on the average (across firms $i = 1, ..., N_t$) pair-wise generalised residual correlation coefficient. Under the null of cross-sectional independence, the CD statistic tends to a standard normal variate. The econometric analysis and the results of these tests are presented in Section 5.1.1.

4.1 Hypothesis testing: signal and noise

We examine a range of possible restrictions on the parameters of equation (7). In section 5.1.2 we explore the significance of individual coefficients and in section 5.1.3 we test the joint hypotheses which we now describe.

For the qualitative survey data to be clearly "useful", or contain a signal about the quantitative data, the MPI turnover growth terms in (7) 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:

$$H^1_0 : \gamma_1 = \gamma_2 = \cdots = \gamma_{12} = 0. \quad (13)$$

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 might expect lags in (7), beyond the horizon of interest (namely three/four months), to be statistically insignificant:

$$H_0^2 : \gamma_{k+1} = 0 \text{ and } \lambda_k^{(j)} = 0, \text{ for all } k > 4,\ j = 0 \text{ and } 2. \tag{14}$$

But rejection of H_0^2 need not imply that firms do not follow the instructions of the CBI by looking further into the past than asked when reporting their trend output growth. When fluctuations in turnover are substantially accounted for by fluctuations in stocks, as appendix A explains, the statistical significance of lagged values for $\Delta x_{i,t}$ may reflect the endogeneity of turnover growth rather than the direct reaction of ITS respondents to lagged $\Delta x_{i,t}$. But irrespective of whether turnover growth is exogenous or not, when firms follow the CBI's instructions we should expect the lagged qualitative responses $y_{i,t}^{(j)}$ ($j = 0$ and $j = 2$) to be statistically insignificant at lags beyond the horizon deemed of interest by the CBI (i.e., three or four months). At shorter lags, say $y_{i,t-1}^{(j)}$, we should expect dependence even when firms follow the CBI's instructions since $y_{i,t}^{*}$ and $y_{i,t-1}^{(j)}$ overlap given that they refer, respectively, to growth over the last three/four months and growth one month ago relative to four/five months ago.

When H_0^2 is rejected we isolate the cause by breaking (14) down into two constituent tests:

$$H_0^3 : \gamma_{k+1} = 0, \text{ for all } k > 4 \tag{15}$$

and

$$H_0^4 : \lambda_k^{(j)} = 0, \text{ for all } k > 4,\ j = 0 \text{ and } 2 \tag{16}$$

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}.

Finally we explore the possibility that the lag in turnover growth should be longer than the eleven months set out in equation (7). We do this as a variable addition test, adding extra terms to equation (7): $\gamma_{13}\Delta x_{i,t-12} + \gamma_{14}\Delta x_{i,t-13} + \gamma_{15}\Delta x_{i,t-14} + \gamma_{16}\Delta x_{i,t-15}$, and testing, relative to the unrestricted equation, H_0^5 where:

$$H_0^5 : \gamma_{k+1} = 0, \text{ for all } k > 12. \tag{17}$$

4.2 A quantitative indicator of firm-level 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 the firm-level quantitative data. This involves inverting the probit models, via Bayes' Theorem (see also Mitchell et al. (2010)), 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 the MPI data at time t, $\Delta x_{i,t}$, given their publication lag. But given the possibility that the ITS data tell us not just about $\Delta x_{i,t}$, but lags of $\Delta x_{i,t}$, we construct indicators of $\Delta x_{i,t-k}$ $(k = 0, 1, 2, ...)$.

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

$$f(\Delta x_{i,t}, \cdots, \Delta x_{i,t-k} | \{\Delta x_{i,\tau}\}_{\tau=1}^{t-k-1}) \sim N(\mu, \Sigma). \tag{18}$$

We need to work out the density function of $\Delta x_{i,t-k}$ conditional on the firms' observed qualitative survey responses at time t and earlier, and conditional on lagged quantitative information $\{\Delta x_{i,\tau}\}_{\tau=1}^{t-k-1}$, and the macro-economic data (Bal_t and x_t) which, to ease notation, we do not condition on explicitly below but take as read. We denote this density function $f(\Delta x_{i,t-k} | \{j_{i,\tau}\}_{\tau=1}^{t}, \{\Delta x_{i,\tau}\}_{\tau=1}^{t-k-1})$. Our indicator, $D_{i,t-k}$ $(k = 0, 1, 2, ...)$, under squared error loss, is then given as:

$$D_{i,t-k} = E(\Delta x_{i,t-k} | \{j_{i,\tau}\}_{\tau=1}^{t}, \{\Delta x_{i,\tau}\}_{\tau=1}^{t-k-1}) \tag{19}$$

$$D_{i,t-k} = \int_{-\infty}^{\infty} \Delta x_{i,t-k} f(\Delta x_{i,t-k} | \{j_{i,\tau}\}_{\tau=1}^{t}, \{\Delta x_{i,\tau}\}_{\tau=1}^{t-k-1}) d\Delta x_{i,t-k}. \tag{20}$$

This is the expectation of firm-level 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 $D_{i,t-k}$ against the autoregressive benchmark indicator $E(\Delta x_{i,t-k} | \{\Delta x_{i,\tau}\}_{\tau=1}^{t-k-1})$.

Bayes' theorem states that

$$f(\Delta x_{i,t-k}|\{j_{i,\tau}\}_{\tau=1}^{t}, \{\Delta x_{i,\tau}\}_{\tau=1}^{t-k-1}) = \frac{P(j_{i,t}, \Delta x_{i,t-k}|\{j_{i,\tau}\}_{\tau=1}^{t-1}, \{\Delta x_{i,\tau}\}_{\tau=1}^{t-k-1})}{P(j_{i,t}|\{j_{i,\tau}\}_{\tau=1}^{t-1}, \{\Delta x_{i,\tau}\}_{\tau=1}^{t-k-1})} \tag{21}$$

where

$$P(j_{i,t}, \Delta x_{i,t-k}|\{j_{i,\tau}\}_{\tau=1}^{t-1}, \{\Delta x_{i,\tau}\}_{\tau=1}^{t-k-1}) = \tag{22}$$

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

and b integrates out the random effect α_i

$$b = \left(\int_{-\infty}^{\infty} P(j_{i,t}|\{j_{i,\tau}\}_{\tau=1}^{t-1}, \{\Delta x_{i,\tau}\}_{\tau=1}^{t}, \alpha_i) f(\alpha_i) d\alpha_i \right). \tag{23}$$

The denominator of (21) involves integrating $\Delta x_{i,t-k}$ out from (22). Note that when $k = 0$, since future values of the quantitative data do not enter the probit models, (22) reduces to

$$P(j_{i,t}, \Delta x_{i,t}|\{j_{i,\tau}\}_{\tau=1}^{t-1}, \{\Delta x_{i,\tau}\}_{\tau=1}^{t-1}) = \int_{-\infty}^{\infty} bf(\Delta x_{i,t}|\{\Delta x_{i,\tau}\}_{\tau=1}^{t-1}) d\Delta x_{i,t}. \tag{24}$$

Given $f(\Delta x_{i,t-k}|\{j_{i,\tau}\}_{\tau=1}^{t}, \{\Delta x_{i,\tau}\}_{\tau=1}^{t-k-1})$, with μ and Σ estimated by least squares, all of the above integrals may be calculated by numerical evaluation.

Estimators $\widehat{P}(j_{i,t}|\{j_{i,\tau}\}_{\tau=1}^{t-1}, \{\Delta x_{i,\tau}\}_{\tau=1}^{t}, \alpha_i)$ for $P(j_{i,t}|\{j_{i,\tau}\}_{\tau=1}^{t-1}, \{\Delta x_{i,\tau}\}_{\tau=1}^{t}, \alpha_i)$ are given by substitution of the estimators $\widehat{\sigma}_{\alpha}^2$, $\hat{\mu}_j$ $(j = 0, ..., 3)$, $\widehat{\gamma}_l$ $(l = 0, ..., 11)$, $\widehat{\lambda}_m^j$ $(m = 1, ..., 12,\ j = 0$ and $2)$, $\widehat{\delta}_1$ and $\widehat{\delta}_2$, in (9). Hence, a feasible empirical Bayes estimator

$$D_{i,t-k} = \widehat{E}(\Delta x_{i,t-k}|\{j_{i,\tau}\}_{\tau=1}^{t}, \{\Delta x_{i,\tau}\}_{\tau=1}^{t-k-1}) \tag{25}$$

may be obtained by numerical evaluation. The impact of the use of plug-in (estimated) parameters, instead of priors for these parameters, is expected to be small when the likelihood dominates the prior for the parameters, which it does in large samples and/or when the priors are vague. Deely & Lindley (1981) show that the empirical Bayes predictor is a first order approximation to the Bayes predictor.

5 Results

5.1 Signal or noise?

5.1.1 Probit Equations

Dynamic ordered probit models, (7) and (8), with $p = 1, ..., 11$ lags of $y_{i,t-1}^{(j)}$ ($j = 0$ and $j = 2$) and $\Delta x_{i,t}$ are estimated on the sub-panel of 96 firms with at least 12 consecutive matched responses. Use of this sub-panel is necessary in order to estimate dynamic models which allow firms' qualitative responses to be potentially affected by events and their own qualitative responses up to a year ago; below in Section 5.1.4 we consider how robust results are to use of this 96 firm sub-panel. The Bayesian Information Criterion (BIC) is then used to select the preferred number of lags, p. This process is different from identifying individually insignificant coefficients, an issue we discuss later, since those associated with very short lags will not be restricted to zero if longer lags play a statistically significant role. Table 4 reports the estimation results for this preferred model. The model is estimated with both the raw data, prior to outlier treatment, and the Winsorised data. The estimated coefficients are closer to zero for the raw data than the Winsorised data; as expected, Winsorisation appears to remove noise from the MPI data and delivers coefficient estimates with higher (robust) t-statistics as well as a model with better overall fit, evidenced by a higher value for the maximised log-likelihood ($\ln L$) and, in turn, a lower value for the BIC in Table 4. This gives a statistical reason for preferring to work with the results from the Winsorised data and henceforth we confine attention to the results from these.

In Table 4 we see that the coefficients on the quantitative data, $\Delta x_{i,t}$, are statistically insignificant at time t and it is lagged values of $\Delta x_{i,t}$ that help explain the qualitative data. The lagged qualitative data also play an important role with larger t-statistics on the coefficients of $y_{i,t-1}^{(j)}$ ($j = 0$ and $j = 2$) than on $\Delta x_{i,t-1}$ and $\Delta x_{i,t-2}$. We cannot reject the restriction that $\sigma_{\alpha}^2 = 0$ and so confine attention to the pooled dynamic ordered probit model. Seasonal, time, sectoral and size dummies were included in the probit model but found to be statistically insignificant. We cannot reject the hypothesis that seasonal and time dummies are jointly insignificant with a p-value of 71%. Sectoral and size dummies are also proved to be jointly insignificant with a p-value of 6.5%. The macro-economic data, Δx_t, were found to be insignificant with a p-value of 21.6% and were excluded from the model. However, the balance statistic, as shown in the table, was statistically significant with a p-value of 0%. The diagnostic statistics in Table 4 also indicate that the generalised residuals are free from non-normality and cross-sectional dependence lending support to our modelling approach.

Table 4: Estimation output for the preferred ordered probit model chosen using the BIC

Explanatory variables	Estimated coeff.		Robust t-stat	
	raw data	Winsorised data	raw data	Winsorised data
$\Delta x_{i,t}$	-0.0254	0.0696	-0.36	-0.43
$\Delta x_{i,t-1}$	0.2070	0.6794	2.36	3.76
$\Delta x_{i,t-2}$	0.2960	0.7590	3.81	4.34
$y^u_{i,t-1}$	1.0177	1.0339	9.43	9.53
$y^d_{i,t-1}$	-0.8639	-0.8440	-7.75	-7.54
$y^u_{i,t-2}$	0.6023	0.6176	5.44	5.55
$y^d_{i,t-2}$	-0.4520	-0.4719	-4.27	-4.43
Bal_t	0.0105	0.0108	2.83	2.93
lower threshold μ_1	-0.8495	-0.8496	-13.01	-13.00
upper threshold μ_2	0.9109	0.9266	14.08	14.15
Number of obs ($\sum_{i=1}^{N} T_i$)			1033	1033
N			96	96
Wald χ^2_8			412.68	421.76
$Prob > \chi^2$			0.00	0.00
Pseudo R^2			0.2631	0.2687
$\ln L$			-811.642	-805.477
BIC			1692.69	1680.36
CD test p-value			0.56	0.64
Normality test p-value			0.24	0.17
p-value: $\sigma^2_\alpha = 0$			0.17	0.24

Note: Estimated using the raw MPI data and MPI data with 5% Winsorisation. Wald χ^2_8 is a Wald test for insignificance of the coefficients on all the explanatory variables and $Prob > \chi^2$ is the associated p-value. $\ln L$ is the maximised value of the log-likelihood; BIC is the Bayesian Information Criterion; CD is the cross-sectional independence test.

5.1.2 Significance of Individual Lags

To provide further indication of the relationship between the ITS and MPI data we now turn to the unrestricted model based on estimation of the dynamic ordered probit models, (7) and (8), with $p = 11$ lags of $y_{i,t-1}^{(0)}$, $y_{i,t-1}^{(2)}$ and $\Delta x_{i,t}$.. In the upper panel of Figure 1 we plot the t- statistics of the estimated coefficients on $\Delta x_{i,t}$ to $\Delta x_{i,t-11}$. The t-statistics do not, of course, necessarily represent the relative importance of the different parameters. However, since the standard errors of the different parameters whose t-statistics are shown in these two graphs are very similar at different lags, the profiles shown in Figure 1 also represent the relative importance of the different lags.

The top panel of Figure 1 shows that the t-values for the short lags of $\Delta x_{i,t-k}$ are often greater than ± 1.96, the 95% critical value. But, as in Table 4, one can reject the view that the qualitative survey data provide a good coincident indicator of growth, since the estimated coefficient on $\Delta x_{i,t-k}$ has a t-value of only -0.4 at $k = 0$.

Figure 1 shows that growth as reported in the firm-level qualitative survey responds most strongly to monthly growth one and two months earlier ($k = 1$ and $k = 2$) and not contemporaneously. The peak influence is 2-3 months ago ($k = 2$), with the signal weakening thereafter. However, the t-values remain greater than 1.96 up to 5-6 months ago ($k = 5$), and are less than 1.96 thereafter except for a spike at $k = 11$. As discussed above this conclusion may be because of the interaction between output and sales rather than because firms look further back than the CBI requests. This finding apparently contradicts the earlier model which, in the light of the BIC was restricted to two lags. However, the BIC is known for its property of leading to parsimonious models and these are, in turn, widely found to be robust in modelling applications; see Clements & Hendry (1998).

The firm-level results are given alongside those from an analogous auto-regressive distributed-lag (ARDL) model estimated on aggregated (macroeconomic) data. ARDL models are used widely in macroeconomics to model dynamic relationships; e.g., see Hendry et al. (1984) and, as mentioned above, Pesaran (1997). The ARDL model is estimated using manufacturing output growth, Δx_t, calculated from the manufacturing component of the IoP (which shows output and not sales) and the balance statistic, Bal_t, from the ITS described in the account of equation (8). 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), the ARDL model is estimated by OLS with the qualitative data (the balance statistic Bal_t) as the regressand. As with (7), again twelve lags of the regressor (Δx_t) and regressand are considered. The balance statistic is available from January 1985 and, after allowing for lags, our estimation period is therefore January 1986 to December 2007. We tested the stability of the regression by estimating over two sub-periods, January 1986 to December 1997 and January 1998 to December 2007; we cannot reject the hypothesis of parameter stability on the basis of the Chow test with $F(25, 239) = 0.86$, $p = 66\%$.

The peak t-statistic in Figure 1 for the macroeconomic data is at $k = 4$, rather than $k = 2$. This indicates that the macroeconomic signal lags the firm-level signal.

We also looked at the t-statistics associated with the estimated coefficients on the lagged dummy variables, $y^{(j)}_{i,t-k}$ $k = 1$ to 12 ($j = 0, 2$). We found that the significance of the lags falls off after $k = 2$, and the pattern is similar for both firms reporting down $j = 0$ and reporting up, $j = 2$. In part, as discussed, this is due to the overlapping nature of the responses rather than genuine state dependence. However with $k = 10$ the coefficient $\lambda^{(2)}_{10}$ had a t-ratio greater than 2. We also examined the analogous t- statistics on the lags of the balance statistic from the macro-economic regression. There is, of course, only one parameter rather than two associated with the lags in this case since the balance statistic replaces the two dummy variables. These t-statistics showed somewhat greater persistence. For a lag of 2 the t- statistic was only 1.6 but it rose above two with a lag of $k = 3$, and was also above two (in absolute value) with lags of $k = 7$ and $k = 9$.

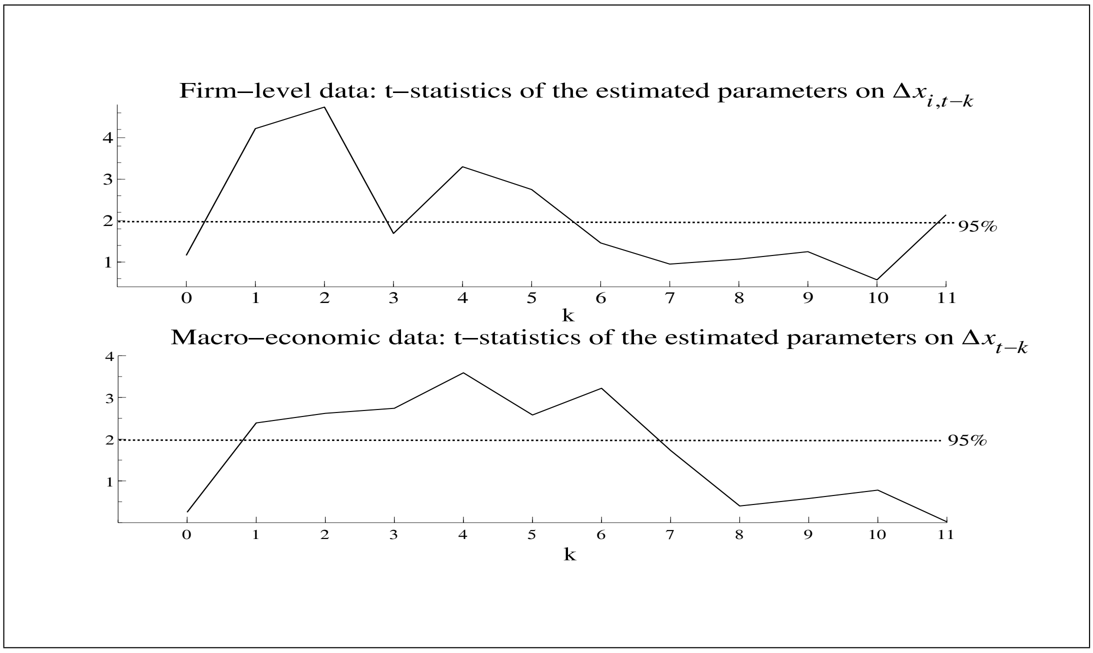

Figure 1: t-statistics for the estimated coefficients on the rolling 1-month MPI turnover growth rates in the firm-level data (top panel) and macroeconomic data (bottom panel). 95% critical values are ±1.96 for both panels

5.1.3 Joint Hypothesis Tests

An alternative perspective is offered by testing the range of hypotheses about joint significance of the coefficients which we discussed in Section 4.1. Table 5 summarises the results of these tests, again based on estimation of the unrestricted dynamic ordered probit models, (7) and (8), with eleven lags.

Table 5: Signal or noise? p-values for the hypothesis tests on the firm-level data and the macroeconomic data

Likelihood ratio tests:	Firm-Level	Macro
H_0^1	0.0001	0.0012
H_0^{2a}	0.1672	0.0003
H_0^{2b}	0.1975	0.0025
H_0^{3a}	0.0417	0.0038
H_0^{3b}	0.0749	0.0524
H_0^{4a}	0.5004	0.0185
H_0^{4b}	0.4777	0.0134
H_0^5	0.4986	0.5673

Notes: $H_0^1 : \gamma_1 = \gamma_2 = \cdots = \gamma_{12} = 0$; $H_0^2 : \gamma_{k+1} = 0$ and $\lambda_k^{(j)} = 0$, for all $k > 4$, $j = 0$ and 2; $H_0^3 : \gamma_{k+1} = 0$, for all $k > 4$; $H_0^4 : \lambda_k^{(j)} = 0$, for all $k > 4$, $j = 0$ and 2; $H_0^5 : \gamma_{k+1} = 0$, for all $k > 12$. Due to the change in the reference period of the ITS question over our sampling period, H_0^{2a}, H_0^{3a} and H_0^{4a} denote the tests for $k > 3$ and H_0^{2b}, H_0^{3b} and H_0^{4b} denote those for $k > 4$. The hypothesis tests are defined fully in Section 4.1.

The p-values in Table 5 indicate that the hypothesis of noise, H_0^1, is clearly rejected with a p-value less than 1% for both the firm-level and macroeconomic data. The ITS data are plainly related to the responses the same firms give to the MPI. But since it is unclear how exactly firms interpret the ITS question, the remaining hypothesis tests in Table 5 shed more light on what the respondents actually had in mind. The failure for the firm-level data to reject H_0^{2a} or H_0^{2b} (with p-values of 17% and 20%) indicates that firms do appear to follow the CBI's instructions 'quite' closely by basing their qualitative responses on MPI growth over the last 3 to 4 months; we discuss this further below. 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%.

Exploration of hypothesis H_0^{3a} again draws attention to a conflict between the results of the hypothesis test and the use of the BIC criterion to specify the equation of Table 4. We can see that we reject the hypothesis that the coefficients on all lags of $\Delta x_{i,t-k}$ longer than three months can be set to zero on the basis of this hypothesis test, with a p-value of 4%.

Nevertheless, comparison of the p-values for H_0^{3a} and H_0^{3b} in Table 5 indicates that the signal from the qualitative data does weaken when related to longer lags of $\Delta x_{i,t-k}$; the p-value rises from 4%, for H_0^{3a} when $k > 3$, to 7% for H_0^{3b} when $k > 4$. But these p-values are lower than those for H_0^{4a} and H_0^{4b}, which are around 50%. This indicates that the aforementioned support for the view that firms look no further than 3 or 4 months into the past (seen in p-values of 17% and 20% for H_0^{2a} or H_0^{2b}) is misleading. Once the qualitative data at longer lags are excluded we find that rejection of the view that firms do not look back further than three/four months is marginal.

In the macro equation we reject, with a p-value of 0.38%, the hypothesis $H_0^{3a} : \gamma_{t-k} = 0$, for all $k > 3$ much more firmly. This implies that the balance statistic, more obviously than the firm-level responses, 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 $\gamma_{t-k} = 0$, for all $k > 4$, with a p-value of 5.24%. Long lags of the balance statistic are significant in the macro equation much more than in the firm-level equation since H_0^{4a} and H_0^{4b} are firmly rejected in the latter but easily not rejected in the former case. Finally there is no suggestion that very long lags of $\Delta x_{i,t}$ or Δx_t are relevant; we cannot reject H_0^5 with p-values of 49.9% with the firm-level data and 56.7% with the macro data.

Thus our overall conclusion is that it is marginal whether one can accept the hypothesis that firms responses to the qualitative survey reflect only the quantitative growth rates that they report within the last three/four months. The outcome of the test depends on the precise nature of the hypothesis tested. But, since statistical significance of the longer lag(s) may be due to the fact that firms report sales rather than output to the MPI, one cannot firmly conclude that firms report output movements over a period longer than the CBI requests.

5.1.4 Robustness of the signal

Only those firms with at least twelve consecutive observations, or thirteen in the level of turnover, are included in the model, (7) and (8). This means our sample dropped from the 807 different firms seen in Table 1 to 96 firms. To provide some check on how reliable results from this sub-panel are we undertook two exercises. Full details are reported in the working paper version of this paper.

First we estimated variants of (7) and (8) that restricted the specification so it is akin to testing the significance of the partial (polyserial) correlation between the ITS and MPI data for a given value of k; see Olsson et al. (1982). The implied restrictions, while rejected by the data (as shown by Table 5) mean that a larger sub-panel of the 807 matched firms can be considered. Importantly, and reassuringly, we found that the main results from Tables 4 and 5 were robust to imposition of these restrictions. The results strengthened our earlier conclusions and led us to conclude that, on balance, there is statistical evidence to indicate that MPI data more than three or four months ago do influence firms when replying to the ITS. We also found that when we estimated analogous models at the sectoral level or for different sized-firms these different firms displayed a similar relationship with the quantitative data to that seen in Figure 1.

Secondly, we undertook tests for sample selection. This involved, following Verbeek & Nijman (1992), adding test variables to the random effects ordered probit model considered in Table 4. Results supported the view that the results from the 96 firms are free from sample selection. There is no tendency for these 96 firms to be either particularly *good* or *bad* at replying to the ITS, although there is obviously a risk that it is lack of power rather than the genuine absence of selection effects which explains our results.

5.2 The ITS as an early indicator of the MPI

We assess the predictive power of the indicator $D_{i,t-k}$ given in (25) 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).

Table 6 summarises the performance of $D_{i,t-k}$ 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 $D_{i,t-k}$ is distinguished according to the prior density, $f(\Delta x_{i,t}, \cdots, \Delta x_{i,t-k} | \{\Delta x_{i,\tau}\}_{\tau=1}^{t-k-1})$, chosen to model the MPI data; for robustness, we consider two choices, an AR for each element of $(\Delta x_{i,t}, \cdots, \Delta x_{i,t-k})'$ with $p = k+1, ..., 11$ lags and an AR with just a single lag. The latter might be expected to benefit from the generally favourable effects of parsimony on forecasting performance; see Clements & Hendry (1998). The final two columns of Table 6 summarise the performance of fitted values from these two models. Since $D_{i,t}$ is the mean of the posterior density, comparison against the prior mean, which is the fitted value generated from the autoregressive model ignoring the survey data, tells us about the value of the qualitative survey data.

While the RMSE of the indicator $D_{i,t-k}$ is, in general, lower than the benchmark, purely autoregressive indicator, the differences in Table 6 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 Tables 4 and 5, the ITS data do contain a signal about the MPI data this signal does not translate into noticeably improved predictive power for the MPI relative to benchmark autoregressive forecasts. Consistent with Figure 1, where it was seen that $\Delta x_{i,t}$ does not help explain firms' contemporaneous qualitative responses, Table 6 shows that what gains there are to using the posterior indicator, which as indicated are minimal, are confined to lags of $\Delta x_{i,t}$. Therefore, in the case of most interest, we find that conditioning our forecast of the latest MPI data, $\Delta x_{i,t}$, on the latest ITS data, which are available ahead of the MPI data, does not deliver more accurate nowcasts.

Table 6: Predictive performance of the ITS indicator of firm-level MPI turnover growth (summarised by its correlation and Root Mean Squared Error)

	$D_{i,t}$: Posterior Mean		Prior Mean	
	AR(11)	AR(1)	AR(11)	AR(1)
Corr with $\Delta x_{i,t}$	0.565	0.402	0.564	0.401
RMSE	0.205	0.228	0.205	0.228
	$D_{i,t-1}$: Posterior Mean		Prior Mean	
	AR(10)	AR(1)	AR(10)	AR(1)
Corr with $\Delta x_{i,t-1}$	0.554	0.417	0.543	0.402
RMSE	0.212	0.231	0.214	0.233
	$D_{i,t-2}$: Posterior Mean		Prior Mean	
	AR(9)	AR(1)	AR(9)	AR(1)
Corr with $\Delta x_{i,t-2}$	0.515	-0.129	0.512	-0.395
RMSE	0.215	0.252	0.215	0.253

$D_{i,t}$ is the mean of the posterior density. The prior mean is from the autoregressive model, which ignores the survey data.

Table 7 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 7 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_{t-p}) and current and

Table 7: Predictive performance of the ITS balance statistic for aggregate manufacturing output growth relative to a benchmark (prior) AR model in manufacturing output growth (summarised by its correlation and Root Mean Squared Error)

	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×100	0.855	0.877	0.857	0.890
	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×100	0.845	0.868	0.857	0.891
	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×100	0.843	0.865	0.858	0.891

$(q-1)$ lagged values of the balance statistic from the ITS. The fact that the conclusions from the firm-level and macro indicators are very similar suggests that the firm-level results are not distorted by the fact that the MPI examines sales while the ITS asks about output volumes. The absolute values of the RMSEs are much lower with the aggregate data than with the firm data for the simple reason that aggregate output is much more stable than is firm-level turnover.

Given the firm-level results in Table 6 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. But a more general point to bear in mind is that autoregressive models tend to perform very well in periods of stability, while they can often be outclassed in more volatile periods. These results show that, in the stable period examined, early estimates of output growth generated using the CBI survey are very little different from those which can be computed without using the qualitative survey.

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 perceived 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 that they should report on output movements over the last three or before July 2003, four months, 'quite' closely with the peak signal two relative to three months ago. However, on balance, the statistical evidence points to the signal remaining statistically significant up to about six months in the past; although it is explained that, given the MPI asks about sales or turnover while the ITS asks about output, this can arise even when firms do follow the CBI's instructions if unexpected fluctuations in sales growth are met from stocks. A clearer conclusion is that the firm-level qualitative data do not provide a good coincident indicator of growth. This was confirmed when, having introduced a novel means of inferring the official quantitative data from the qualitative data, we found that conditioning autoregressive forecasts of the MPI data on contemporaneous values of the ITS data does not improve inference more than trivially. This suggests that the CBI survey has little role to play in enhancing our knowledge of what has recently happened to manufacturing output, a result which is confirmed when we examine the signal generated by macro-economic data.

Nevertheless, the finding that the responses to the ITS are related to the quantitative returns provided by the same firms in the MPI does suggest that confidence can be placed in the responses to questions which have no counterpart in official enquiries. In the monthly survey these include questions about order books for domestic and export sales and levels of stocks. In the wider quarterly survey there are also questions covering, among other things, business optimism, capital expenditure plans, capacity utilisation, factors likely to limit output and expenditure on training. While verification of these is obviously desirable, the results we find in the first part of our study provide no reason to doubt that the ITS offers valid indicators of the business environment.

7 Acknowledgements

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 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 three anonymous referees, an associate editor, the joint editor (Antony Fielding) and participants in presentations at the National Statistics Methodology Advisory Committee (NSMAC), the CBI, the Bank of England and the MMF (Birkbeck, 2008), RES (Surrey, 2009) and ISAE (Rome, 2009) conferences. We gratefully acknowledge financial support from the ESRC (Award Reference: RES-062-23-0239) and the Bank of England. This work contains statistical data from ONS which are Crown copyright and reproduced with the permission of the controller of HMSO and Queen's Printer for Scotland. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates. For more details about the CBI's ITS, see http://www.cbi.org.uk/ndbs/content.nsf/802737AED3E3420580256706005390AE/2F172E85D0508CEA80256E20003E95C6.

A Appendix: A two-equation model for output growth and sales growth

The general model, (7), as long as a sufficient number of lagged terms in $\Delta x_{i,t}$ are included, can be motivated as the solution of a two-equation model, which accommodates the potential endogeneity of sales growth. For a related model see Smith & Blundell (1986); see also the discussion in Pesaran (1997) which shows, in the context of the ARDL models introduced above for the macroeconomic data, the importance of augmenting the lag order to accommodate the potential endogeneity of the explanatory variables. In our context, endogeneity could arise from the measurement error in relating output growth to sales growth. Let $s^*_{i,t}$ represent sales growth, where $s^*_{i,t}$ is a linear function of $\Delta x_{i,t}$ and up to $l^* < 11$ lags, such that $s^*_{i,t} = f(\Delta x_{i,t}, \Delta x_{i,t-1}, ..., \Delta x_{i,t-l^*})$. Then consider the two-equation system:

$$y^*_{i,t} = \varkappa_i s^*_{i,t} + \alpha_i + u^1_{i,t} \quad \text{(A-1)}$$

$$s^*_{i,t} = \pi_i s^*_{i,t-1} + u^2_{i,t} \quad \text{(A-2)}$$

where $|\pi_i| < 1$ to ensure stationarity of sales growth, and in the first equation, (A-1), output growth, $y^*_{i,t}$, relates to sales growth, $s^*_{i,t}$, plus an error term, $u^1_{i,t}$, which might in part capture the change in stock movements. For notational ease only we suppress dependence of $y^*_{i,t}$ on the lagged dummies, $y^{(j)}_{i,t}$. The second equation, (A-2), then assumes that sales growth follows an autoregressive process, that is taken to be first order. Additional lagged terms can be included if necessary. $(u^1_{i,t}, u^2_{i,t})$ are assumed to be mean zero and jointly normally distributed random variables with variances σ_{11} and σ_{22}, respectively, and covariance σ_{12}. Under these assumptions $u^1_{i,t} = (\sigma_{12}/\sigma_{22})u^2_{i,t} + \epsilon_{i,t}$, where $\epsilon_{i,t}$ is a mean zero disturbance distributed independently of $u^2_{i,t}$. When $\sigma_{12} = 0$ fluctuations in sales growth are not met from stocks and lead directly to output movements. When $\sigma_{12} \neq 0$, $s^*_{i,t}$ is endogenous (correlated with $u^1_{i,t}$) and the derivation of the relationship between $y^*_{i,t}$ and $s^*_{i,t}$ should allow for this contemporaneous feedback (or indirect relationship). This is achieved by substituting in (A-1) which is then seen, by augmenting the lag order of (A-1), to generate a (conditional) model of the type seen in (7). The coefficients on $\Delta x_{i,t}$, and its lags, in (7) can then be seen to be a non-linear function of $\varkappa_i$, $(\sigma_{12}/\sigma_{22})$, π_i and the assumed functional form for $f(\Delta x_{i,t}, \Delta x_{i,t-1}, ..., \Delta x_{i,t-l^*})$. So while we might expect σ_{12} to be negative, if unexpected fluctuations in sales growth (shocks to $u^2_{i,t}$) are met from stocks, rather than increases or decreases in output, we do not have a prior view on the sign of the coefficients on $\Delta x_{i,t}$, and its lags, in (7).

References

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

Butler, J. & Moffit, R. (1982), 'A computationally efficient quadrature procedure for the one-factor multinomial probit model', *Econometrica* **50**, 761–764.

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

Deely, J. L. & Lindley, D. V. (1981), 'Bayes empirical Bayes', *Journal of the American Statistical Association* **76**, 833–841.

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.

Hendry, D. F., Pagan, A. R. & Sargan, J. (1984), Dynamic specification, *in* Z. Griliches & M. D. Intriligator, eds, 'Handbook of Econometrics'.

Hsiao, C., Pesaran, M. H. & Pick, A. (2007), Diagnostic tests of cross section independence for nonlinear panel data models. Cambridge Working Papers in Economics No.0716.

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

Lui, S., Mitchell, J. & Weale, M. R. (2009), Collective sentiment in qualitative business survey. National Institute of Economic and Social Research Discussion Paper No.328.

Machin, S. J. & Stewart, M. B. (1990), 'Unions and the financial performance of British private sector establishments', *Journal of Applied Econometrics* **5**, 327–350.

Mitchell, J., Smith, R. J. & Weale, M. R. (2002), 'Quantification of qualitative firm-level survey data', *Economic Journal* **112**, C117–C135.

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

Mitchell, J., Smith, R. J. & Weale, M. R. (2010), 'Efficient aggregation of panel qualitative survey data'. National Institute of Economic and Social Research Discussion Paper No.261 (Revised)).

Ohlsson, E. (1995), Co-ordination of samples using permanent random numbers, *in* B. Cox, D. Binder, B. Chinnapa, A. Christianson, M. Colledge & P. Kott, eds, 'Business Survey Methods', Wiley, New York, pp. 153–169.

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

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

ONS (2007), General information on the Quarterly Inventory (Stocks) Inquiry, note = Available at http://www.statistics.gov.uk/StatBase/Source.asp?vlnk=1444&More=Y.

ONS (2009), A Brief Guide to the IDBR. http://www.statistics.gov.uk/idbr/idbr.asp.

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. H. (1997), 'The role of economic theory in modelling the long run', *Economic Journal* **107**(440), 178–91.

Pesaran, M. H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica* **74**(4), 967–1012.

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.

Smith, R. J. & Blundell, R. W. (1986), 'An exogeneity test for a simultaneous equation Tobit model with an application to labor supply', *Econometrica* **54**, 679–685.

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