

“The impact of survey aggregation methods on the
quality of business survey indicators”
ECFIN/2003/A3-04. Final Report.

James Mitchell, Kostas Mouratidis and Martin Weale*
National Institute of Economic and Social Research, London

November 8, 2004

Contents

1	Executive Summary	3
2	Introduction	6
2.1	Plan of this report	7
2.2	The disaggregate approach for drawing inferences about official data from survey responses: an overview	7
2.3	Aggregate indicators: an overview	10
2.4	Examination of the prospective survey responses	11
2.5	Out-of-sample performance	12
3	The Data	12
3.1	Relating the survey data to the official data	12
3.1.1	Choice of the reference series	14
3.2	Summary statistics about the surveys	15
4	Comparing the in-sample performance of the retrospective aggregate and disaggregate indicators of manufacturing output growth	15
4.1	The performance of the retrospective disaggregate indicators as a function of the cut-off value	16
4.2	Illustrative results for alternative definitions of the reference series	16

*Address for correspondence: James Mitchell, 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: j.mitchell@niesr.ac.uk. We are very grateful to DG-ECFIN for helpful comments
and suggestions on interim and draft final versions of this report.

4.3	Detailed results for specific cut-off values	23
4.4	Intuition - beginning to understand why the disaggregate indicators work better in-sample	29
5	Comparing the out-of-sample performance of the retrospective aggregate and disaggregate indicators of manufacturing output growth	30
6	Comparing the in-sample performance of the prospective aggregate and disaggregate indicators of manufacturing output growth	33
7	Comparing the out-of-sample performance of the prospective aggregate and disaggregate indicators of manufacturing output growth	34
8	References	46
A	Disaggregate indicator: technical details about firm-level quantification	49
A.1	Ordered Discrete Choice Models	49
A.2	Specification tests	50
A.3	Inferring the Official Data	51
B	Aggregate Quantification Techniques: A Review	54
B.1	The Probability Approach	54
B.1.1	Carlson and Parkin's Method	55
B.2	The Regression Approach	56
B.2.1	Relating the Regression Approach to the Probability Approach . . .	56
B.3	The Reverse-Regression Approach	57
B.3.1	Motivating the Regression Formulation	58
B.3.2	Estimation of x_t	59
B.3.3	Relating the Reverse-Regression Approach to the Probability Approach	59
C	The disaggregate indicator using prospective survey data	61
C.1	Firm-Level Quantification of Prospective Survey Data	61
C.2	The relationship between the prospective survey data and the official data	62
C.3	Inferring the Official Data in the Prospective Case	63
C.4	Producing out-of-sample forecasts from the qualitative survey data	66
D	Detailed tables of results for the retrospective survey data: in-sample	67

1 Executive Summary

1. This study applies, and extends, the recently developed techniques of Mitchell, Smith and Weale (2002a) for drawing inferences about official data from survey responses by exploiting the underlying panel data set of individual survey responses. This has led to the production of so-called disaggregate survey indicators, that are based on relating individual survey responses to the official data.
2. From the Joint Harmonised EU Programme of Business and Consumer Surveys firm-level responses are considered for four European countries: Germany, Portugal, Sweden and the UK.
3. The survey data for Germany, Portugal and Sweden are available monthly, while for the UK they are available quarterly. Therefore the relation between the survey data and official data for manufacturing output growth is considered at a monthly frequency for Germany, Portugal and Sweden and at a quarterly frequency for the UK. For Germany, Portugal and Sweden the official data considered are the 3-monthly growth rate of manufacturing output at a quarterly rate; for the UK quarterly growth is considered. Although our focus remains on these reference series, alternative definitions are considered for completeness. The sample period for Germany is 199101-200012 (120 monthly observations), for Portugal 199406-200312 (115 monthly observations), for Sweden 199601-200312 (96 monthly observations) and for the UK 1988q3-1999q3 (45 quarterly observations).
4. Both the retrospective or *ex post* (output trends over the past three months) and prospective or *ex ante* (output trends over the next three months) survey questions are considered.
5. Consideration of the prospective responses required us to extend the theoretical framework of Mitchell, Smith and Weale (2002a) that considered the retrospective responses only. This project shows how the prospective or forward looking individual responses can be combined to produce forecasts of output growth - we consider how the qualitative prospective survey data published at time $(t - 1)$ can be converted into a quantitative indicator of expected growth in period t . This theoretical development constitutes an important theoretical output of this project not foreseen at its outset. Both parametric and nonparametric versions of the disaggregate indicator are derived; in practice we largely confine attention to the nonparametric indicator in the prospective case. This is because the nonparametric indicator is both easier to implement and because experimentation, carried out in detail for the U.K. as part of this project, suggests that it works better than the parametric indicator.
6. Exploiting these firm-level survey responses disaggregate indicators are derived. We have related firm's retrospective and prospective opinions about output from the survey to official data for manufacturing output growth. Across all four countries

this new method of aggregating survey data delivers improved signals of economic activity in-sample; the disaggregate indicators offer more information about manufacturing output growth than traditional so-called aggregate survey indicators that relate aggregated survey data to official data; see Table 1 that considers the retrospective and prospective survey questions.

Table 1: In-sample correlation of the aggregate and disaggregate indicator against manufacturing output growth: summary results

	Aggregate			Disaggregate	
	Pes	CP	CSW	Parametric	Nonparametric
Retrospective Case					
Germany	0.412	0.416	0.412	0.675-0.770	0.685-0.782
Portugal	0.077	0.100	0.121	0.341-0.758	0.475-0.754
Sweden	0.215	0.225	0.232	0.471-0.685	0.635-0.740
UK	0.567	0.586	0.603	0.621-0.872	0.666-0.921
Prospective Case					
Germany	0.287	0.272	0.248		0.667-0.745
Portugal	0.166	0.032	0.170		0.510-0.703
Sweden	0.213	0.208	0.230		0.638-0.794
UK	0.596	0.540	0.586		0.794-0.897

Notes: Pes denotes Pesaran; CP denotes Carlson-Parkin; CSW denotes Cunningham, Smith and Weale. The disaggregate indicator is computed for a range of values of what is called the cut-off parameter. In each case the correlation with official output growth data is computed and the range of results is presented.

7. We supplement this in-sample analysis with a comparison of the performance of the proposed disaggregate indicators with traditional aggregate indicators in simulated out-of-sample experiments designed to mimic real-time application of the procedures. Table 2 compares the forecasting accuracy of the aggregate and disaggregate forecasts in the retrospective and prospective cases. They show that in Germany and particularly the UK the disaggregate forecasts are more accurate than their aggregate counterparts.
8. The disaggregate indicators are based on aggregation across firms. Experimentation with alternative means of weighting the firm-level series revealed that, in general, using equal weights delivers the best results.
9. More work is required to consider whether the improved in-sample performance of the disaggregate indicators, relative to the aggregate indicator, can be translated into better performance out-of-sample across all four countries. Both the manner

in which this indicator is re-scaled, and the way in which the firm-level quantified series are weighted, should be central to this.

Table 2: RMSE of the aggregate and disaggregate indicators against manufacturing output growth in the out-of-sample period: summary results

	Aggregate			Disaggregate	
	Pes	CP	CSW	Parametric	Nonparametric
Retrospective Case					
Germany	4.125	4.012	4.038	3.596-3.863	3.534-4.014
Portugal	9.382	9.250	9.232	9.744-12.591	10.479-13.019
Sweden	6.490	6.489	6.403	6.352-6.683	6.793-7.071
UK	4.255	3.898	3.854	2.699-3.946	2.418-4.137
Prospective Case					
Germany	4.647	4.437	4.524		4.255-4.317
Portugal	9.440	9.316	9.298		8.691-10.297
Sweden	7.064	6.985	6.926		7.508-9.247
UK	3.906	3.763	3.292		2.089-3.035

Notes: RMSE denotes root mean squared error; Pes denotes Pesaran; CP denotes Carlson-Parkin; CSW denotes Cunningham, Smith and Weale. The disaggregate indicator is computed for a range of values of what is called the cut-off parameter. In each case the RMSE against the subsequent outturn for official output growth data is computed and the range of results is presented.

2 Introduction

The use of qualitative survey data, such as those released by DG-ECFIN, as a complement to official data continues to be popular. Official data are published at a considerable lag. Policy-makers, *inter alia*, are therefore keen to rely on data from a variety of sources, such as surveys, in order to form an impression of the state of the economy before “hard” figures are published.¹ Furthermore, surveys usually ask not only about experiences in the recent past but also about prospects for the near future; they may therefore be helpful in providing a guide to short-term prospects in a way that official data cannot.

This type of survey typically involves ordered qualitative responses, namely respondents answer “up”, “the same” or “down”. Moreover these data are traditionally available in aggregate form rather than at the individual level. Published survey data typically report only the proportion of respondents who report “up”, “the same” or “down”, or the balance of opinion (“up”’s minus “down”’s).² There have been numerous studies of the way in which these survey responses link to and anticipate official data for both output and price movements. There are two main approaches for linking the survey data to official data and deriving quantitative estimates of actual and expected output or price movements, the *probability method* of Carlson and Parkin (1975), and the *regression method* of Pesaran (1984). The two approaches are used widely to obtain quantitative estimates of actual and expected output or price movements.

However, the interpretation and analysis of such data is usually based on aggregation of individual responses in a way which may lose information contained in the pattern of individual responses and does not exploit the panel aspect of the data. Little attention to date has been paid to whether better signals of economic behaviour can be derived from analysis of the panel data set of individual or firm-level responses underlying the aggregate responses, or proportions.³ One exception is by Mitchell, Smith and Weale (2002a) who for the U.K. consider how the retrospective survey responses of individual firms can be combined if the aim is to produce an early indication of official output data, based on the fact that survey data are published ahead of official data on output growth. They find that more accurate indicators, so-called “disaggregate” indicators, are obtained when quantification proceeds in a manner which allows for a degree of heterogeneity across firms.

This project provides a comprehensive examination of the performance of the disaggregate indicators. This involves the following three developments: (i) the consideration of firm-level data for Germany, Portugal and Sweden, as well as the UK; (ii) examination of the out-of-sample performance of the various indicators across the four countries and (iii) consideration of how the prospective or forward looking individual responses can be

¹Of course most official data are also compiled from quantitative surveys. Here we use the term survey to refer to those surveys which ask for qualitative answers to questions.

²There is often a small number of “don’t knows”. Analysis of such surveys is usually done looking only at the respondents who do know.

³There has been limited previous work using individual responses to surveys [see Nerlove, 1983; Horvath, Nerlove and Willson, 1992; McIntosh, Schiantarelli and Low, 1989; Branch, 2004; Souleles, 2004]. But this work has focused on testing the nature of expectation formation.

combined to produce forecasts of output growth.

2.1 Plan of this report

The plan of the remainder of this report is as follows. Sections 2.2 and 2.3 provide non-technical descriptions of the disaggregate and aggregate approaches for drawing inferences about official data from retrospective survey responses; further technical details are provided in Appendices A and B. Section 2.4 considers how forecasts of output growth can be derived from the prospective survey data. Section 2.5 considers evaluation of the aggregate and disaggregate indicators based on their out-of-sample performance. Section 3 then summarises some properties of the survey data and the official data for manufacturing output growth; it considers both the survey data themselves and the choice of which reference series for manufacturing output growth to use. Sections 4, 5, 6 and 7 then compare the in-sample performance of the aggregate and disaggregate indicators in the retrospective case, the out-of-sample performance of the indicators in the retrospective case, the in-sample performance using the prospective survey data and the out-of-sample performance using the prospective data, respectively.

Figures 3-6 summarise the in-sample performance of the aggregate and disaggregate indicators for Germany, Portugal, Sweden and the UK in the retrospective case. Figures 9-12 summarise the out-of-sample performance of the aggregate and disaggregate indicators for Germany, Portugal, Sweden and the UK in the retrospective case. Figures 13-16 summarise the in-sample performance of the aggregate and disaggregate indicators for Germany, Portugal, Sweden and the UK in the prospective case. Figures 17-20 summarise the out-of-sample performance of the aggregate and disaggregate indicators for Germany, Portugal, Sweden and the UK in the prospective case. Further tables of results for each of the four countries are provided in Appendix D.

2.2 The disaggregate approach for drawing inferences about official data from survey responses: an overview

The disaggregate approach, originally proposed by Mitchell, Smith and Weale (2002a) for examination of the retrospective survey responses and evaluated only in-sample for the U.K., comprises two steps. At the first step for each respondent in the survey the time-series of individual (respondent-level) categorical responses are converted into quantitative series for the aggregate or macroeconomic variable of interest, say economy or region-wide output growth. Then at the second step the disaggregate indicator of the aggregate variable is derived by averaging (either unweighted or weighted) this quantitative series across respondents at a given point in time.⁴

⁴An alternative so-called “semi-disaggregate” approach, proposed by Mitchell, Smith and Weale (2002b), examines whether the link between survey responses and published data can be improved on by looking at respondents’ responses in the light of the responses they had made in the previous period. The approach is semi-disaggregate in the sense that although based on respondent specific information the final estimating equations are not respondent specific.

Let us now consider each of these steps in a little more detail. At the first step, a time-series of survey responses for a given individual is related to official data on the variable of interest, say output growth. The relationship is motivated by assuming that respondents' survey responses are triggered by an unobserved continuous random variable as it crosses thresholds. An individual's response is assumed to relate linearly to the aggregate (economy-wide) quantitative variable of interest. Using an ordered discrete choice model we can estimate for each respondent its relationship with the observed aggregate data. Specifically, we assume the disturbance to this equation has a logistic cumulative distribution function and fit ordered logistic models. Alternatively, and without loss of generality, a cumulative normal distribution could be assumed instead and probit models fitted.

At the second step, having estimated, say, the ordered logistic model separately for each respondent we infer the (quantitative) value of the aggregate variable from the survey data. As survey data are usually published ahead of the official data this provides early quantitative estimates. This is achieved from the estimated models by calculating the most likely, or expected, value for the aggregate "hard" variable given an individual's categorical response at a given point in time.

The expected value for the macroeconomic variable given an individual's categorical response is calculated in two ways. The first route exploits Bayes' Theorem. It uses numerical integration to obtain the probability density function (pdf) of the aggregate variable conditional on an individual's response and then derives the expected value of this conditional pdf. To calculate this conditional expectation we use the estimated relationship between respondent's survey responses and the aggregate variable and also make a parametric assumption about the pdf governing the aggregate variable. We considered two pdf's, the normal and the Pearson. The Pearson family of density functions provides approximations to a wide variety of observed distributions and can offer a more general characterisation of a sample than the normal distribution, allowing for skewness and kurtosis. However, in practice results using the Pearson distribution were little different to using the normal distribution and attention below is confined to the normal distribution.

An alternative non-parametric route adopted is, rather than assuming a parametric structure for the conditional probability of output growth given an individual's survey response, to evaluate this probability by taking the mean of the empirical distribution function. This amounts to a simple, intuitive and transparent approach that should be attractive to potential users of the output of this project. Take the average (across time) of the macroeconomic variable when the individual replied "up", "the same" and "down". Then the quantitative series for this individual involves replacing an "up" with the average when the individual replied "up", and so for "the same" and "down". Note that this approach is operational in-sample. We consider out-of-sample modifications below.

The disaggregate indicator of the official macroeconomic variable, the expected value for the macroeconomic variable, is then defined as the average across respondents (either unweighted or weighted) of their expected values for the macroeconomic variable at a given point in time. The nature of the weighting depends on our view of the sampling

process. Since we have no *a priori* view, we will consider both unweighted and weighted averages. We compare their performance in the empirical applications considered below; see Sections 4, 5, 6 and 7.

Unweighted averages may be viewed as appropriate since they can be seen to filter out those respondents that offer little information about the aggregate variable of interest. On the other hand, we will consider various weighting schemes. These will include using those (what we might consider theoretical) weights, measuring “firm-size”, supplied by the survey-data provider and used by them when aggregating to obtain the weighted proportion of firms that replied “up” and “down”. We will also consider various empirical weights, such as weighting the individual series in inverse proportion to the variance of the disturbance to the discrete choice model. Since this is not estimated, being fixed *a priori*, it is proxied by some measure of the fit of the estimated discrete choice model- we consider the pseudo R^2 . Alternatively, we examine the role the “polyserial” correlation coefficient can play in this respect; see Olsson, Drasgow and Dorans (1982).⁵ This coefficient offers a nonparametric measure of association between ordered and continuous variables and can therefore be used to correlate (across time) the ordered survey responses for a given respondent with the continuous variable of interest, such as aggregate output growth or inflation. Since the survey data are ordered, the Pearson correlation between the survey and official data can offer a biased measure of association since it does not account for the discrete nature of the survey data; see Coenders, Satorra and Saris (1997). The polyserial correlation coefficient assumes the variables have a bivariate normal distribution, and can be therefore seen to be imposing an additional distributional assumption compared with logit/probit estimation where no distributional assumptions are made about the independent variable.

Finally, we consider weighting according to the fit of the relationship between the respondent-level (quantified) series itself and the macroeconomic variable of interest. This could be based on the correlation or root mean squared error (RMSE) between them. That respondent-level series most correlated (or with the lowest RMSE) will receive the largest weight, and so on. For those individuals whose responses are orthogonal to the macroeconomic variable of interest, their quantitative series will equal the mean of the macroeconomic variable. Therefore, if there are a large number of individuals whose responses have no information for the macroeconomic variable then the disaggregate indicator either unweighted or weighted will itself be squeezed towards the mean. We will, therefore, consider a number of ways of correcting this. One is to ensure that low or zero weights are given to those individuals whose responses contain no information. Another, used by Mitchell, Smith and Weale (2002a), is to use a regression equation *post hoc* to align the disaggregate indicator to the aggregate (official) series.

The proposed disaggregate indicators require sufficient time-series observations for a given respondent to ensure the parameters in the discrete-choice models are consistently

⁵The polyserial correlation coefficient is reported widely by software packages, such as LISREL, and used by practitioners as a descriptive statistic for ordered and categorical data [see Nerlove, 1988]. Significance can be assessed using the asymptotic standard errors, derived from the Hessian matrix evaluated at the maximum likelihood.

estimated. Similarly, in the non-parametric case, consistency of the sample moments requires sufficient time-series observations. Assuming the sample observations are independent these observations need not be consecutive. There will be no induced bias. Individuals that do not respond enough are dropped from the sample used to derive the disaggregate indicators.

There are no hard-and-fast rules guiding the choice of how many observations per respondent are necessary. Below in Sections 4, 5, 6 and 7 we accordingly take an eclectic approach and when examining the performance of the disaggregate indicators consider a range of values. These values are the minimum number of time-series observations per firm necessary for its inclusion in the disaggregate indicator; a particular firm may well respond to more surveys but if it responds to fewer it is dropped.⁶

Since firms that do not respond to at least these number of surveys are dropped from the sample used to derive disaggregate indicators of manufacturing output growth, there is a danger that the sample selection could induce bias in the disaggregate indicators.⁷ In any case, notwithstanding the implied theoretical properties of the disaggregate indicators, their value when analysing survey data is determined by how well they perform in practice, both in-sample and out-of-sample, relative to the traditionally used aggregate indicators. This should serve as the main test of their value.

A possible alternative to the disaggregate approach considered here, that avoids the need to drop data for some firms, is to pool the data by imposing homogeneity restrictions across firms and then exploit traditional panel-data estimators; see Hsiao (2003). With a common slope parameter, panel-data estimation techniques have been developed for the binary discrete-choice case. However, T must be small for *fixed*-effects estimation by conditional ML or *random*-effects estimation where T -dimensional integrals need to be evaluated although this difficulty may be alleviated somewhat by the use of simulation techniques for the evaluation of multivariate integrals [namely, the Geweke-Hajivassiliou-Keane simulator]; e.g. see Keane (1994).⁸ However, we do not follow this approach here since our results indicate considerable heterogeneity across firms in their slope coefficients; therefore imposing a common slope coefficient would result in heterogeneity bias.

2.3 Aggregate indicators: an overview

The traditional approach to converting the findings of surveys into quantitative estimates of movements in economic activity has been to take the aggregate findings of such surveys—the proportion of firms reporting that output has risen, stayed the same or fallen— and relate them to official output data. Approaches suggested have included the balance

⁶In any case, we should not recommend estimation of ordered logit/probit models with fewer than about 20 time-series observations if the goal is reliable parameter estimation (since consistency of the estimators in the firm-level models is predicated on $T \rightarrow \infty$), rather than examination of the performance of the parametric disaggregate indicator irrespective of its theoretical properties. Of course, this choice too is arbitrary and warrants further investigation *via* Monte-Carlo experiments.

⁷We therefore do consider a test for sample selection in Section 4.3.

⁸These methods are not directly applicable for the trichotomous ordered model considered here, where T is small for some firms but large for others.

statistic, the probability method [Carlson and Parkin, 1975], the regression method [Pesaran, 1984] and the reverse regression method [Cunningham *et al.*, 1998]. These methods are formally reviewed in Appendix B. For an example of an application using Carlson and Parkin’s approach, see Lee (1994). The regression approach is used, for example, by the Bank of England [see Britton, Cutler and Wardlow, 1999]. One could also consider other aggregate quantification techniques; there have been numerous extensions of the probability and regression approaches - e.g. see Smith and McAleer (1995). Our experience suggests that similar results would be obtained using the traditional probability and regression approaches. Therefore we focus on the traditional interpretation of these approaches; they provide a natural benchmark against which to compare the performance of the proposed disaggregate indicators.

We consider the aggregate indicators both when the aggregate proportions are calculated using equal weights and when they are calculated with firms weighted according to their size.

2.4 Examination of the prospective survey responses

Qualitative surveys usually ask not only about experiences in the recent past but also about prospects for the near future; they may therefore be helpful in providing a guide to short-term prospects in a way that official data cannot. In this section we discuss a means of producing a forecast of official data using the disaggregated responses to a qualitative survey. This required the development of the appropriate methodology.

Specifically, we consider a survey that asks a sample of N_t manufacturing firms near the end of period $(t - 1)$ whether their output growth is expected to rise, not change or fall over period t compared to period $t - 1$. Our method of quantification of the prospective survey responses follows Mitchell, Smith and Weale (2002a) in postulating an underlying relationship between firm specific output growth and the official data for aggregate output growth and then extends their approach to the case of forward-looking or prospective survey responses.

While the nonparametric disaggregate indicator is readily extended from the retrospective to prospective case, the theory behind the parametric disaggregate indicator requires serious modification to accommodate the fact that the use of forward-looking variables induces endogeneity of output growth in the ordered logit models. We have derived an appropriate estimator and we refer the reader to Appendix C for details. However, with the exception of the UK, we do not consider the use of the parametric disaggregate indicator in the applications below since this estimator is computationally quite difficult to deal with, and in any case for the UK results indicated that especially out-of-sample the nonparametric method delivers better results.

The reference series in the prospective case is assumed to be the same as that in the retrospective case, but is considered at a lead reflecting the forward looking nature of the prospective survey question. Specifically, for the UK the prospective survey responses published in quarter t are related to quarterly output growth in quarter $(t + 1)$ relative to quarter t . For Germany, Portugal and Sweden the prospective survey responses published

in month t are related to the three monthly growth of output in month $(t + 3)$ relative to month t .

2.5 Out-of-sample performance

It is well known that improved in-sample fit need not translate into better out-of-sample performance; e.g. see Clements and Hendry (1999). It is therefore important to study the performance of the alternative survey based indicators in simulated out-of-sample experiments. We will divide our sample into two periods, an estimation and an evaluation (or out-of-sample) period of two or three years. The estimation period is the initial sample period used to estimate the alternative indicators of economic activity. Then based on recursive estimation of the indicators in the out-of-sample period, an experiment designed to mimic real-time computation of the indicators, we will compare the performance of the disaggregate and aggregate indicators. To extract the quantitative respondent-level series in real-time from the panel of survey responses, the time t value should exploit survey data up to time t , but macroeconomic information only up to time $(t - k)$, where $k \geq 1$, at least when the macroeconomic variable of interest is GDP, or some other output index, as these data are typically published with a lag of at least a quarter.

One possibility will be to tune the weighting scheme in the light of the performance of the indicators in the previous recursive sample. For example, we will consider weighting the respondents at time period t , $t = 1, \dots, t$, according to how well correlated the respondent-level series was with the aggregate variable of interest using data up to time $(t - k)$. This allows the weighting pattern to change over the out-of-sample period. Interesting questions can be raised about whether, for example, there are individuals whose responses are always more informative about the macroeconomic data, or whether there is considerable variation across time in terms of the ‘best’ individuals.

3 The Data

3.1 Relating the survey data to the official data

The retrospective and prospective survey data on output (where firms have been asked to reply “up”, “the same” or “down”) are related to official data for the growth rate of manufacturing output. The retrospective question provides the basis of deriving timely indicators of manufacturing output growth, x_t , given that the survey data are published more rapidly than official data, while the prospective question provides the basis for deriving forecasts of output growth.

The survey data for Germany, Portugal and Sweden are available monthly, while for the UK they are available quarterly.⁹ Therefore the relation between the survey data

⁹The UK exercise has been restricted to the quarterly frequency for the following reason. The Confederation of British Industry (CBI) charged for supplying the data. There was no provision in this project to pay for the data at a monthly frequency. From a previous research grant (used as the basis for our earlier work on panel survey data) we already had quarterly data from 1988q3-1997q3. For this project

and official data for manufacturing output growth is considered at a monthly frequency for Germany, Portugal and Sweden and at a quarterly frequency for the UK. For Germany, Portugal and Sweden the official data considered are the 3-monthly growth rate of manufacturing output at a quarterly rate; for the UK quarterly growth is considered. The sample period for Germany is 199101-200012 (120 monthly observations), for Portugal 199406-200312 (115 monthly observations), for Sweden 199601-200312 (96 monthly observations) and for the UK 1988q3-1999q3 (45 quarterly observations).

Figures 1 and 2 plot these official data for output growth.¹⁰ Growth using the monthly data is, as expected, more volatile than the quarterly data. Future work might consider whether the survey-based indicators provide a better signal of some smoothed measure of economic activity than the actual growth rates, such as some measure of the “business cycle”. Focus here, however, remains on the actual data, which are of primary interest to users. For Germany, Portugal and Sweden, where official data for output growth are exploited at a monthly frequency we do, however, consider the use of alternative reference series to the 3-monthly growth rate.

As indicated, our focus is on reference series that are seasonally adjusted. We relate the survey responses to the reference series without making any corrections for any seasonality that might be present in the survey data. That is, no seasonal corrections are made to either the individual-level survey responses or their aggregated counterparts, namely the proportions (of up’s and down’s) and the balance statistic. This is because we wish to treat the individual-level and aggregated survey data similarly. In contrast to the aggregated data, seasonal adjustment of the individual-level data is not straightforward since the data are qualitative. However, we should note that commonly when examining aggregated survey data (such as the balance statistic) users do routinely make seasonal corrections. As seasonal adjustment delivers a smoother series, adjustment tends to deliver an aggregate indicator with a higher correlation against the reference series than when seasonal adjustment is not performed.

Therefore open methodological issues remain concerning seasonality and survey data. Given the qualitative nature of the survey responses, it is not obvious how seasonality might manifest itself. Indeed, certainly in the case of the UK survey, respondents are asked to ignore the likely impact of seasonal factors when replying to the survey: the Confederation of British Industry ask: “Excluding seasonal variations, what has been the trend [is the expected trend] over the past [next] four months with regard to volume of output?”. Future work should explore the relation between survey data and seasonality further. With enough data one interesting experiment would be to quantify using both the aggregate and disaggregate approaches for each month or quarter separately. This

we paid the CBI to update these quarterly data, although the cost was not included in the budget for this project. In fact, it only proved possible for the CBI to supply us with data up to 1999q3 as at the end of 1999 the CBI moved to a new survey processing platform that involved changing the participant identification numbers. This means it is no longer straightforward to match firms pre and post December 1999 which is necessary to construct the panel data set of survey responses.

¹⁰All data were downloaded from *Datastream*. For Germany, Portugal and Sweden we consider monthly data for real manufacturing output growth seasonally adjusted (called INDUSTRIAL PRODUCTION - MANUFACTURING VOLA in *Datastream*).

would facilitate statistical testing for seasonality.

3.1.1 Choice of the reference series

As mentioned above, to some extent the emphasis in this report on the 3-monthly growth rate of manufacturing output may appear arbitrary. Our primary focus on the 3-monthly growth rate (for Germany, Portugal and Sweden) is motivated not just by the fact the survey questions refer to a time horizon of (approximately) three months but by the following argument. If we were to consider, say, the 12-monthly growth rate as the reference series, then presumably when quantifying the survey data at time t official data for the reference series have been published up to say at least $(t - 3)$. This means that at time t hard data on the reference (official) series are in fact known for at least 9 of the 12 months to which the reference series relates. There is the fear, therefore, that we are using the survey data to try and tell us in large part what we already know. Our focus on 3-month growth is motivated by the view that at time t data beyond $(t - 3)$, in other words older than $(t - 3)$, are certainly available (published); 3-monthly growth seems the appropriate reference series for assessing the true informational content of the survey data.

This argument is consistent with earlier work we have undertaken that suggested that the correlation between the retrospective aggregate indicator, based on the traditional balance statistic, and quarterly growth is at its maximum with a lag of about 2 to 3 quarters. We feel that this and the fact that annual growth rates are in any case smoother than quarterly growth rates explains the higher correlation typically found between the retrospective survey data and annual output growth than three-monthly growth. However, official monthly data exist for most of the period covered by the annual growth rate and these in themselves have a high correlation with the annual rate. Thus we feel a comparison of the survey with the immediate past is more informative about the utility of the survey.

For completeness in Section 4.2 we do, however, present both some representative results for two alternative definitions of the reference series, based on the 12-monthly growth rate and the rolling quarter on quarter growth rate, and show that the correlation between the balance statistic and output growth depends on the period of the output index under consideration.¹¹

We also mention here the possibility, not considered in this project as we believe it lies outside its agreed scope, that the individual survey responses be related to sectoral output growth. Certainly previous work we have carried out for the UK indicates that the disaggregate approach also works well at the sectoral level [see Mitchell, Smith and Weale, 2002a]. Perhaps aggregating survey-based indicators/forecasts of sectoral growth will deliver improved estimates of aggregate growth? This is an interesting question for future research.

¹¹Quarter on quarter growth at time t is defined in the level of output y_t as $100 \cdot \ln \left(\frac{y_t + y_{t-1} + y_{t-2}}{y_{t-3} + y_{t-4} + y_{t-5}} \right)$.

3.2 Summary statistics about the surveys

- For Germany the sample records the survey responses of, in total, 9703 firms over the period 1991*m*1 to 2000*m*12 (120 months). There are, on average, only 3843 firms in the sample at time t , with 48 time-series observations per firm
- For Portugal the sample records the survey responses of, in total, 1528 firms over the period 1994*m*6 to 2003*m*12 (115 months). There are, on average, only 832 firms in the sample at time t , with 63 time-series observations per firm
- For Sweden the sample records the survey responses of, in total, 1620 firms over the period 1996*m*1 to 2003*m*12 (96 months). There are, on average, only 784 firms in the sample at time t , with 46 time-series observations per firm.
- For the UK the sample records the survey responses of, in total, 5519 firms over the period 1988*q*3 to 1999*q*3 (45 quarters). There are, on average, only 1142 firms in the sample at time t , with 9.3 time-series observations per firm.

4 Comparing the in-sample performance of the retrospective aggregate and disaggregate indicators of manufacturing output growth

This section considers in some detail the in-sample performance of the retrospective aggregate and disaggregate indicators of manufacturing output growth. Sections 5, 6 and 7 then turn to the out-of-sample performance of the retrospective indicators, the in-sample performance of the prospective forecasts and the out-of-sample performance of the prospective forecasts, respectively.

As indicated above, the disaggregate indicators require sufficient time-series observations for a given respondent to ensure parameters/moments are consistently estimated. We take an eclectic approach and consider the performance of the indicators as a function of the minimum number of observations assumed to be satisfactory; we call this number the “cut-off”. Since this approach involves the consideration of a large number of variants of the disaggregate indicator it is not practical to present detailed results for each cut-off. Accordingly, we present our results by dividing them into two sections. Section 4.1 examines the performance of the unweighted disaggregate indicator as a function of the cut-off. It summarises the performance of the indicators in easy-to-read graphs. But as we shall see in Section 4.3 when more detailed results are presented for specific cut-off values, in general, the unweighted disaggregate indicator performed well relative to various weighting schemes.

4.1 The performance of the retrospective disaggregate indicators as a function of the cut-off value

The performance of the proposed disaggregate indicators of economic activity is compared with traditionally used aggregate indicators by examining the accuracy of the estimates against the outturn (i.e. the official data). The correlation coefficient (Corr.) against the outturn will be reported; in the out-of-sample experiments root mean squared error (RMSE) is considered.

Correlation informs us about the informational content of the indicator series; when the square of the correlation statistic is strictly positive the indicator series explains some of the variation in manufacturing output growth about its mean.¹² A high value for corr. indicates that a strong signal about the outturn may be recovered from the indicator regardless of how the indicator has been scaled, and whether the RMSE is high or low.

Figures 3-6 plot the correlation of the aggregate and disaggregate indicators against the outturn for output growth as a function of the cut-off value. In the lower panel of each figure a histogram indicates the proportion of firms in the total sample present at each cut-off value; as seen in Germany there were in total 9703 firms sampled, in Portugal 1528 firms sampled, in Sweden 1620 firms sampled and in the U.K. 5519 firms sampled.

The findings from Figures 3-6 are striking; across the four countries the disaggregate indicators offer a stronger signal about the official data than the aggregate indicators, irrespective of the cut-off value.

Figures 3-6 also shows that the correlation of the parametric disaggregate indicators against the outturn declines as more observations per firm are considered (i.e. as the cut-off value increases). The fewer firms are dropped the better the fit of the parametric disaggregate indicator. The nonparametric disaggregate indicator achieves its highest correlation against the official data when at least 88 observations are considered in Germany, 99 are considered in Portugal and 70 in Sweden. In the UK the nonparametric disaggregate indicator performs better the lower the cut-off value.

The finding that the explanatory power of the disaggregate indicators is often better the lower the cut-off value, at least in part, is a consequence of over-fitting. A simple example illustrates this. Consider the case where there are just T firms in the sample and they each reply once but at different points in time. In this case the nonparametric disaggregate indicator will fit the official data perfectly. Given the dangers of over-fitting, it is therefore important to consider the behaviour of the indicators on an out-of-sample basis.

4.2 Illustrative results for alternative definitions of the reference series

In this section we continue to consider the retrospective survey data and provide two illustrations of the sensitivity of inference to the choice of the reference series and conclude

¹²Equivalently, the indicator series has some informational content when, in a linear regression of the outturn on the indicator and an intercept, R^2 is greater than zero.

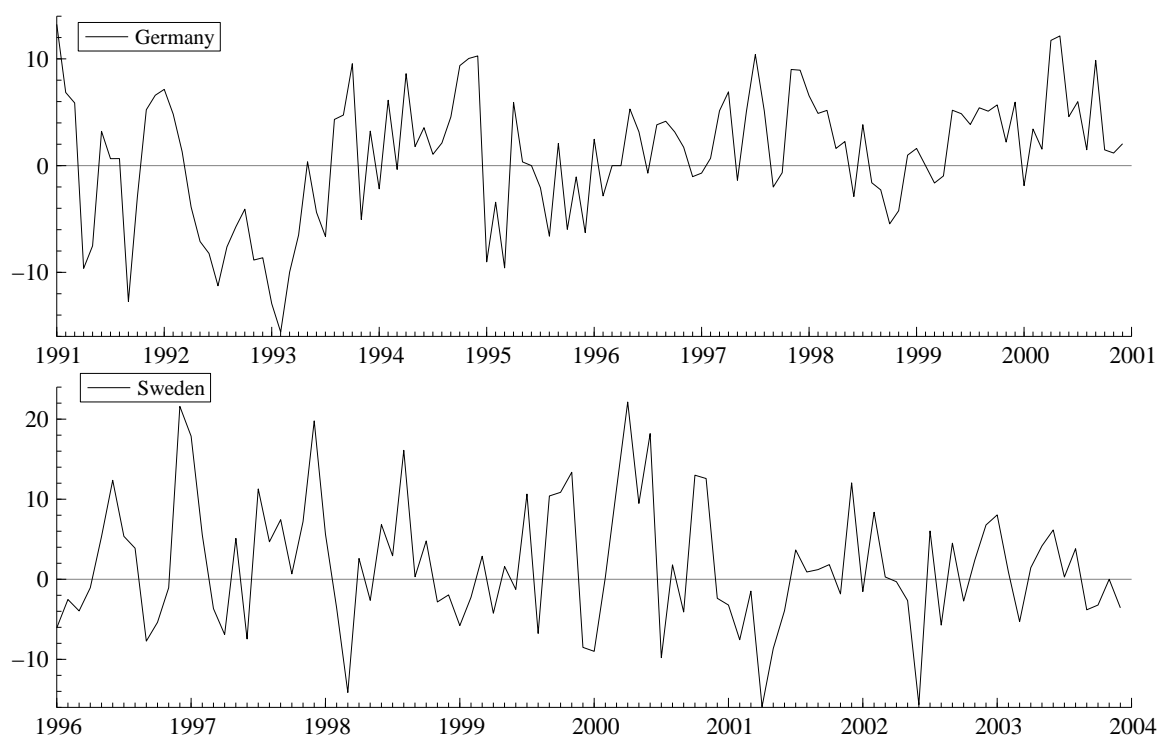


Figure 1: Manufacturing output growth in Germany and Sweden: 3 monthly growth at a quarterly rate (in percentage points)

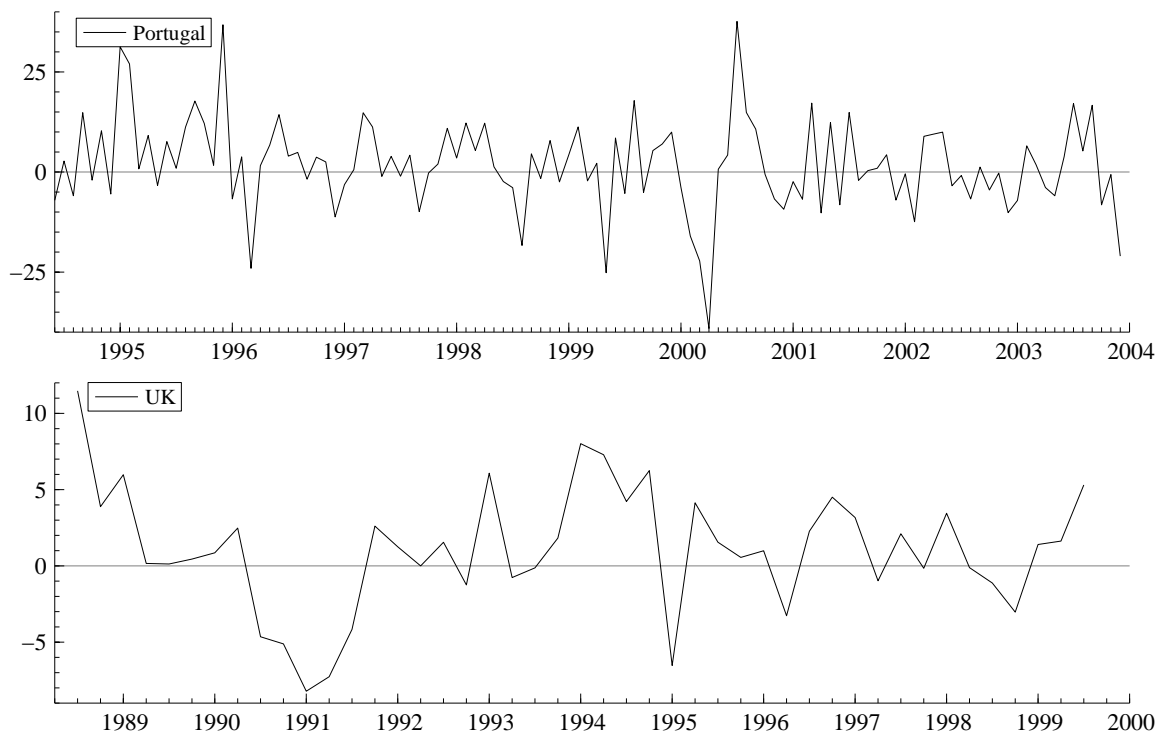


Figure 2: Manufacturing output growth in Portugal and the UK: 3 monthly growth at a quarterly rate (in percentage points) for Portugal and quarterly growth (in percentage points) for UK

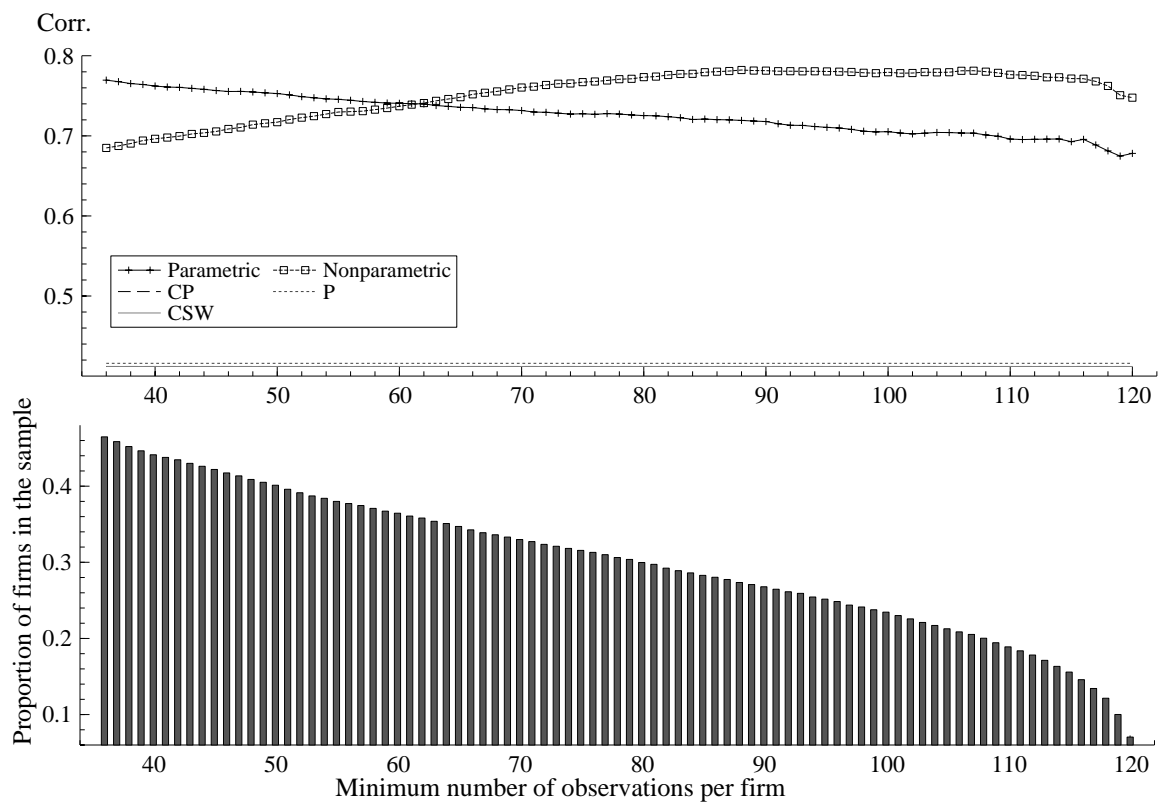


Figure 3: Germany. In-sample performance of the aggregate and disaggregate indicators in the retrospective case. Correlation of the indicators against manufacturing output growth as a function of the minimum number of observations considered per firm (the cut-off)

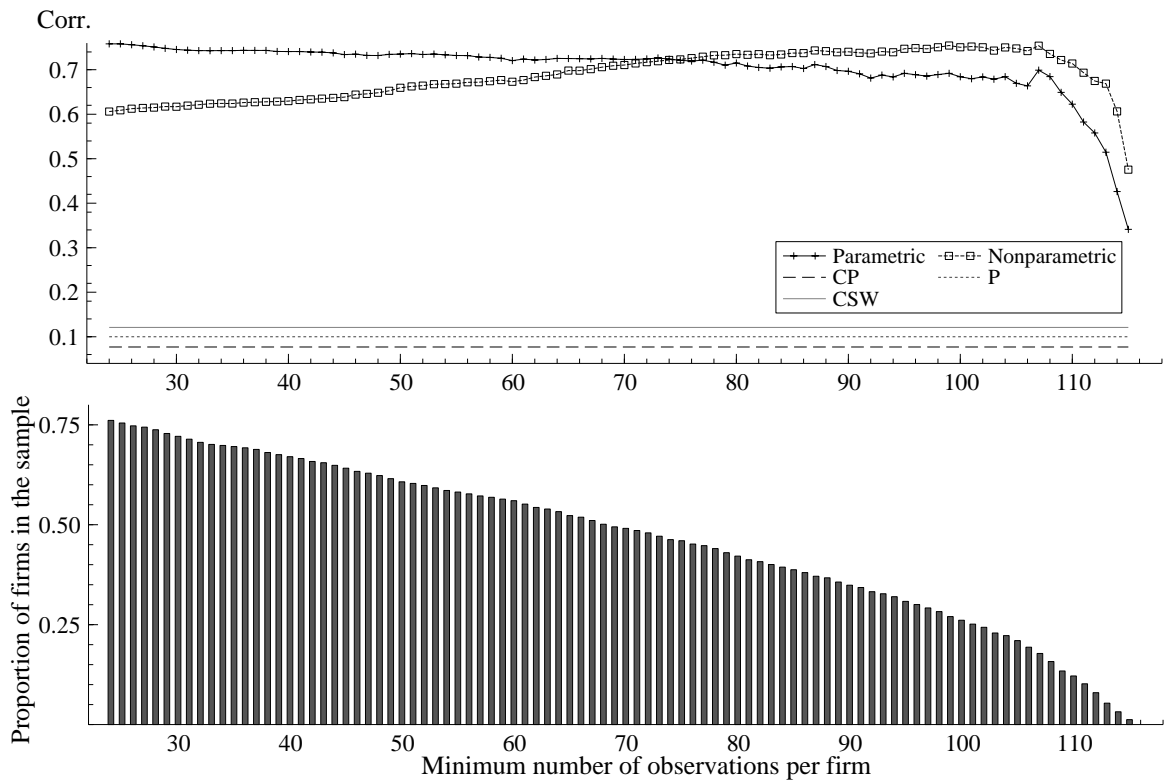


Figure 4: Portugal. In-sample performance of the aggregate and disaggregate indicators in the retrospective case. Correlation of the indicators against manufacturing output growth as a function of the minimum number of observations considered per firm (the cut-off)

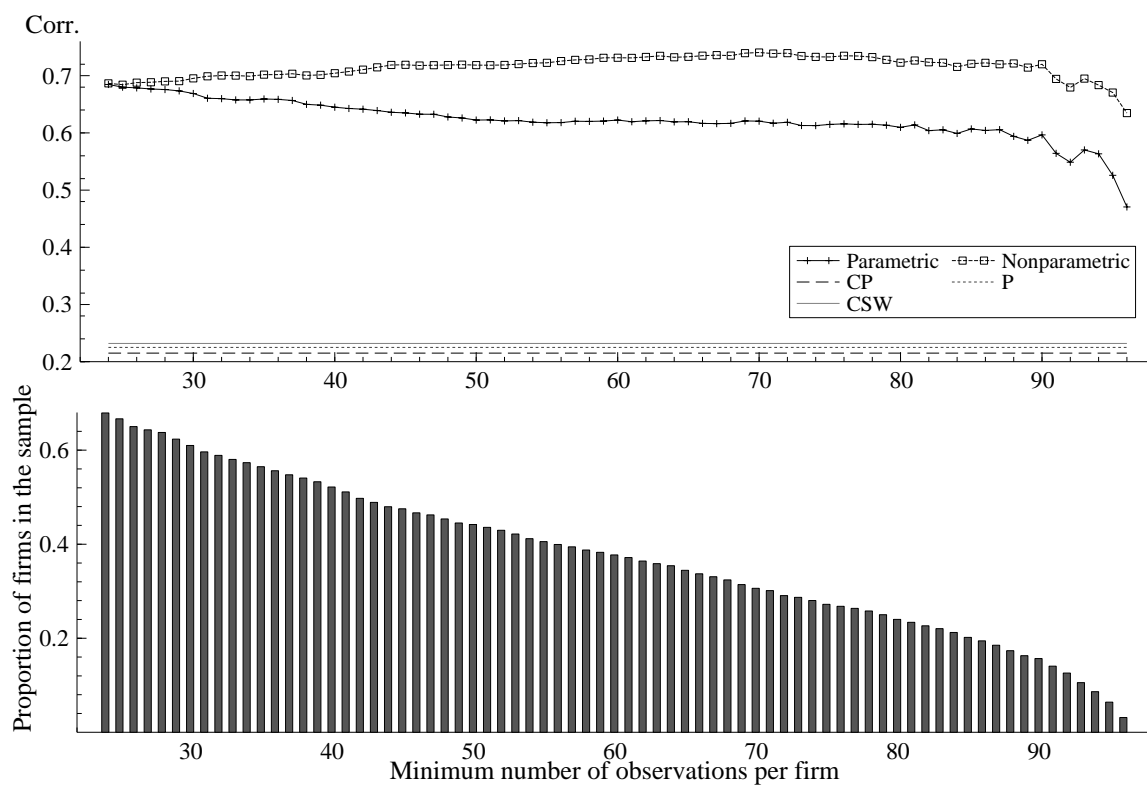


Figure 5: Sweden. In-sample performance of the aggregate and disaggregate indicators in the retrospective case. Correlation of the indicators against manufacturing output growth as a function of the minimum number of observations considered per firm (the cut-off)

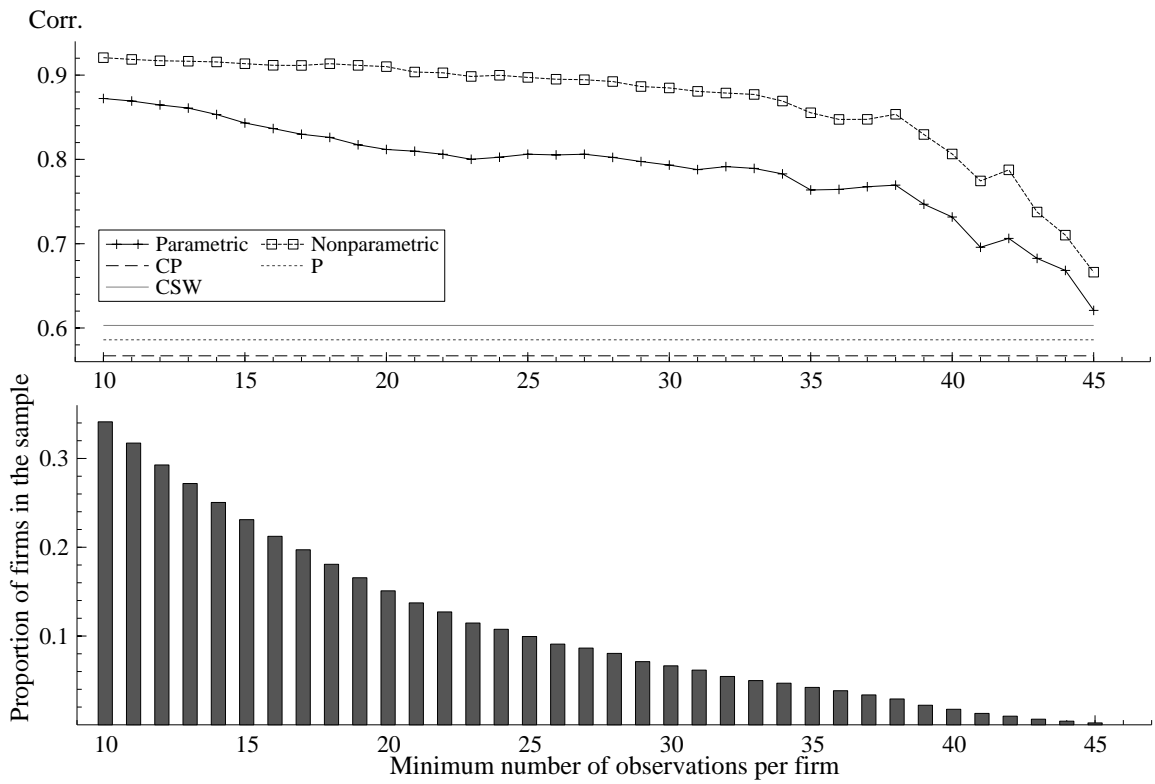


Figure 6: UK. In-sample performance of the aggregate and disaggregate indicators in the retrospective case. Correlation of the indicators against manufacturing output growth as a function of the minimum number of observations considered per firm (the cut-off)

by noting that this sensitivity is to be expected. Firstly, for Portugal we consider the correlation of the aggregate and disaggregate indicators with three alternative definitions of the reference series: (i) three-monthly growth; (ii) twelve-monthly growth and (iii) quarter-on-quarter growth. Correlation is highest against the twelve-monthly growth rate. The Pesaran aggregate indicator, for example, has correlation of 0.1 against the 3-monthly growth rate, correlation of 0.16 against quarter on quarter growth and correlation of 0.217 against the 12-monthly growth rate. The nonparametric disaggregate indicator is also best correlated against 12-monthly growth; see Figure 7.

Secondly, following Weale (2004), we simply show that correlation between the reported retrospective balance statistic, as published by the European Commission, and the growth in the UK manufacturing output index depends on the period of the output index under consideration. As Weale writes, Figure 8 plots “the correlation between the reported balance and the growth in the manufacturing output index over periods from one to twelve months ending in the month of the publication of the index for the period January 1992-February 2004. It can be 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. When we consider growth over the past twelve months the correlation rises to 0.56. It is, however, mistaken to infer from this that the survey is a good coincident indicator because the growth rate over the past twelve months is better regarded as an indicator of the state of the economy six months ago than as an indicator of what is going on at the moment”.

Henceforth, focus is on the three-monthly growth rate as the reference series.

4.3 Detailed results for specific cut-off values

For specific cut-off values we now present further results, against the three-monthly growth rate. Specifically, for each of the four countries, we provide tables that indicate:

1. The performance of the aggregate indicators in not just the full-sample but what are called the included and excluded samples. As indicated above in deriving the disaggregate indicators there is, in a theoretical sense although in practice the barometer of success for the disaggregate indicators is their performance relative to the aggregate indicators, a danger of sample selection since those firms that reply to fewer observations than the cut-off are dropped. Let “included sample” denote those firms with more than, say, 20 time series observations. Let “excluded sample” denote those firms in the full-sample omitted from the included sample. In the absence of sample selection, the included sample may be regarded as a random sample from the full-sample and inference from both included and excluded samples should be equivalent apart from sampling error. That is, indicators or statistics derived from both included and excluded samples should not differ significantly. We therefore consider the correlation of the three traditional aggregate indicators with the out-turn for output growth. To test statistically for sample selection we test:

$$H_0 : r_i - r_j = 0, \tag{1}$$

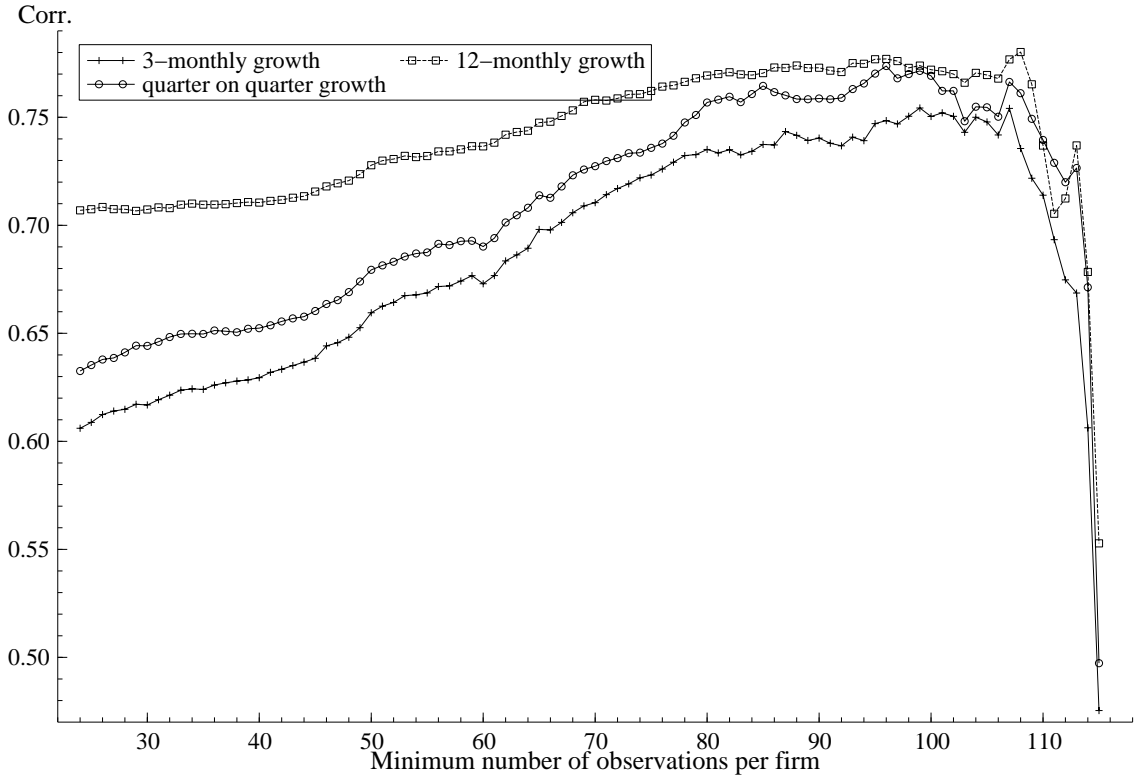


Figure 7: Portugal. In-sample performance of the nonparametric disaggregate indicator for three alternative definitions of the reference series. Correlation of the indicators against manufacturing output growth as a function of the minimum number of observations considered per firm (the cut-off)

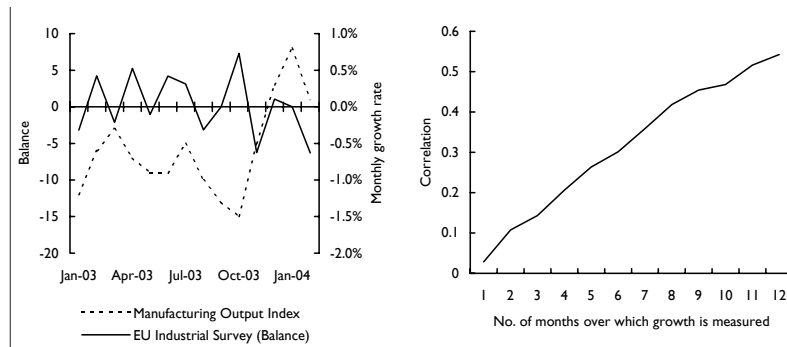


Figure 8: UK output growth and the EU business survey plotted alongside correlations between output growth and the business survey

where r_i (r_j) refers to the correlation coefficient for a given aggregate indicator in the included (excluded) sample. We test H_0 under the assumption that r_i and r_j are independently and normally distributed.¹³ We also indicate how well the aggregate indicators in the full-sample perform when the aggregate proportions (reporting “up” and “down”) are computed by weighting firm’s responses according to their size; see Appendix B for further discussion. These size-weights were given to us by the survey provider. Weights were not available for Sweden. For the UK they are only available up to 1997q3, rather than 1999q3.

2. The performance of the disaggregate indicators subject to alternative means of weighting the firm-level series. Firm-level series using both the parametric and nonparametric approach are weighted using not just equal weights, as above, but also using the following five weighting schemes:
 - (a) Each firm’s (quantified) series is weighted in proportion to how well the logistic model, underlying the parametric disaggregate indicator, fits the data. We proxy the fit of this model by the pseudo R^2 from the estimated ordered logit models. Specifically, we use McFadden’s R^2 ; see Amemiya (1981), p. 1505.
 - (b) Polyserial correlation; see Olsson, Drasgow and Dorans (1982). Rather than weighting each firm’s (quantified) series by the pseudo R^2 , we consider weighting them according to how well the firm’s categorical responses in the survey are correlated with official output growth. Since the polyserial correlation coefficient offers a nonparametric measure of association between ordered and continuous variables it can be used to correlate (across time) the ordered survey responses for a given respondent with the continuous variable of interest, namely aggregate output growth.
 - (c) RMSE; each firm’s (quantified) series is weighted in inverse proportion to its RMSE against aggregate (official) output growth.
 - (d) Correlation; each firm’s (quantified) series is weighted in proportion to how well it is correlated with aggregate (official) output growth.
 - (e) “Firm-size”. These are the weights explicitly provided by the survey provider that are presumably used by them when computing the weighted proportion of firms that reply “up” and “down”. The weights are typically based on firm’s sales volumes. As already mentioned, weights were not available for Sweden. For the UK they are only available up to 1997q3, rather than 1999q3.

We also consider computation of the nonparametric indicator with any missing values for a given firm replaced by that firm’s modal response. Intuitively we can see the consequences of replacing the missing values with the modal response by considering the following scenario: as the number of missing values for a given firm tends to T (the dimension of the time-series) their quantitative series will tend to

¹³Under these assumptions $Var(r_i - r_j) = (1/T)(1 - r_i^2)^2 + (1/T)(1 - r_j^2)^2$.

the mean of the macroeconomic variable. So this approach of dealing with missing values has the effect of forcing the correlation of a firm's quantified series with the macroeconomic variable to fall as the number of missing values rises.

3. The empirical support for the specification underlying the parametric disaggregate indicator. The firm-level ordered logit models were subjected to a series of specification tests. These tests examine whether the assumed linear relation between the firm's survey response and contemporaneous output growth is supported empirically. They examine whether one should allow for dynamic dependence in the official data. It is important to test this given that macroeconomic data are widely accepted to exhibit dependence. Two tests are considered; see Appendix A for details. Both (i) a joint test of no dynamics in terms of how the official data relate to the official data and (weak) exogeneity of the official data in the ordered logit models plus (ii) a score test of mis-specification are considered. This latter test is a joint test for omitted variables (specifically x_{t-1} and, like a RESET-type test for incorrect functional form, powers of $\hat{\beta}_i x_t$), neglected heteroscedasticity and asymmetry of the logistic c.d.f.; see Murphy (1996). Results are presented using both traditional and Bonferroni corrected critical values. We report the proportion of times, across the firms, the specification tests were not rejected.¹⁴ We also present tables indicating the proportion of times (across firms) that the official data are insignificant statistically in the ordered logit models estimated for each firm.
4. The relative performance of alternative survey based indicators of economic activity. Tests for forecasting encompassing are carried out to test the relative informational context of the aggregate and disaggregate indicators; see Harvey, Leybourne and Newbold (1998).

The cut-off values considered for each country are as follows: Germany, 96; Portugal, 80; Sweden; 72 and UK, 20. These values are a compromise between those values where the performance of the disaggregate indicator is best and having a sufficiently long time-series to be reasonably confident that parameter/moment estimation is reliable.

The results, presented for each country in turn in Appendix D, see Tables 6-33, can be summarised as follows:

1. In all cases, see Tables 7, 14, 21 and 28 there was no evidence of a statistically significant difference between the performance of these aggregate indicators in the included and excluded samples. This is not surprising looking at Tables 6, 13, 20 and 27 where we see that the correlation of the aggregate indicators with output growth is similar across the included, excluded and full samples. These results are consistent with the view that the included sample may be regarded as a random sample, and that inference from it should be unbiased. This implies that since the

¹⁴Although results for a given country are presented for just one cut-off value, similar results were obtained for other (similar) cut-off values. Naturally, we should expect the power of these specification tests to decline as fewer observations are considered per firm (i.e. as the cut-off value is reduced).

disaggregate indicators outperform traditional aggregate indicators we can conclude that this improvement is due to disaggregation *per se*, and is not the consequence of using a different sample. We also supplemented the above by using forecast encompassing tests to examine whether the aggregate indicators derived from the excluded firms add information *vis-à-vis* the disaggregate indicators. Again, there was little evidence to suggest that dropping firms led to an informational loss.

2. The aggregate indicators in Portugal, see Table 13, perform particularly poorly compared with their performance in other countries; they offer little explanatory power over output growth. As seen in Figure 7 we did experiment with alternative definitions of output growth in Portugal based on the 12-monthly, rather than 3-monthly, growth rate of manufacturing output. As expected, a stronger signal about the official data was then recovered: correlation is in general about 0.2-0.3, rather than around 0.1 as it is in Table 13. Importantly, the disaggregate indicator also offers a strong signal about the 12-monthly growth rate: correlation is again higher than 0.7. Therefore, our main result is unaffected. It is also interesting that only in Portugal are improved aggregate indicators obtained when the proportions are calculated using weights reflecting the size of the firm.
3. Tables 8, 15, 22 and 29 present detailed results examining the performance of the disaggregate indicators. These tables consider alternative means of weighting the firms in deriving the disaggregate indicators. The tables evaluate the performance of the disaggregate indicators by looking at their mean, standard deviation, and their correlation against official aggregate output growth. We turn to RMSE and “S.D. post regression” below.
 - (a) Looking first at the results for the unweighted (or in fact equal weighted) parametric and nonparametric disaggregate indicators (rows N and NP) we see that despite their sample mean approximately estimating that of the outturn for official output growth, x_t , correctly, they appear too smooth; the standard deviation is low compared to that for the outturn. These disaggregate indicators display too little volatility compared with the outturn and would perform badly using RMSE criteria. This feature of the indicators has been observed elsewhere with alternative indicators [see, for example, Cunningham (1997)]. Less volatility is observed because the scale is incorrect.
 - (b) One explanation for this is based on those firms whose responses are poorly correlated with actual output growth. In the extreme case where responses are uncorrelated with output, the inclusion of these reduces the standard deviation of the indicator but does not affect its correlation with output growth. Excess smoothness of the disaggregate indicators can then be explained by the presence of firms in the sample of survey responses whose responses contain no signal about output growth and are essentially ‘noise’. To reconcile this incompatibility in volatility between outturn and indicators for manufacturing output growth, note that the outturn is the signal recovered from the survey

data plus a residual error component. Rescaling the indicators through linear regression on the outturn is one simple method of obtaining an indicator which tracks output growth as closely as possible. In fact, the RMSE results in Tables 8, 15, 22 and 29 are based on having used a regression equation *post hoc* to align the disaggregate indicator to the aggregate (official) series. The effects of this regression are taken into account in our subsequent out-of-sample analysis. Tables 8, 15, 22 and 29 also report, in the column titled “S.D. post regression”, the standard deviation of the indicator after it has been rescaled by the regression.

- (c) Tables 8, 15, 22 and 29 also consider an alternative based on defining some empirically based metric, e.g. rows R2, poly, RMSE or cor in the Tables, and using that to identify the noisy firms that could then be excluded, or given a lower weight, when defining the disaggregate indicator. We see that use of these alternative weights does help increase the volatility of the disaggregate indicators (the standard deviation increases) but the correlation against the outturn is on occasion reduced. This suggests that use of the regression equation *post hoc* is the preferred means of correcting the scale of the disaggregate indicators since of course the correlation is then not altered. These tables also consider, except for Sweden, the use of weights based on the size of the firm; without exception the disaggregate indicators perform better in the unweighted case. Weighting firms according to their size, as measured by the survey provider, does not deliver an improved signal from the disaggregate indicator.
4. Tables 9, 16, 23 and 30 report the proportion of times, across firms, that there is no evidence for misspecification in the ordered logit models used as the basis for deriving the parametric disaggregate indicator. Two types of test are considered: (i) the joint test for no dynamics and exogeneity and (ii) the score or LM tests. The former test is always very supportive of the specification chosen, particularly when the Bonferroni correction is used - the proportion rises to unity. We know that use of traditional critical values inflates the Type I error. Using the LM test the results are less supportive of the specification chosen, but using the Bonferroni correction the proportion of cases where there is support for the specification encouragingly remains high. There is no evidence of misspecification in Germany for over 80% of firms, in Portugal for 40%-60% of firms, in Sweden for over 90% of firms and in the UK for over 95% of firms. Overall, therefore, the results from the two tests appear to provide empirical support for the use of the firm-level models, (A.2), as the basis of the parametric disaggregate indicators.
5. Figures 3-6 clearly indicate that the disaggregate indicators offer more in-sample information about the official data than the aggregate indicators; they are far better correlated with official output growth. But, it important to supplement this with an explicit statistical test of whether the disaggregate indicators offer more information about official output growth than the aggregate indicators; indeed we can

test whether they completely encompass them. These tests are based on the regression based approach for forecast combination of Granger and Ramanathan (1984) and Harvey, Leybourne and Newbold (1998). We ran OLS regressions of the form: $x_t = a_1 \widehat{disagg}_t + a_2 \widehat{agg}_t$ where x_t is actual growth, \widehat{disagg}_t is growth implied by the disaggregate indicator and \widehat{agg}_t is growth implied by the aggregate indicator.¹⁵ The estimated coefficients in these regressions, \widehat{a}_1 and \widehat{a}_2 , show how the information from the disaggregate and aggregate indicators should be combined to provide the best (in a mean squared error sense) estimates of output growth: \widehat{a}_1 and \widehat{a}_2 tell us the weights that should be attached to the disaggregate and aggregate indicators, respectively. Results from these regressions, that compare the parametric and non-parametric disaggregate indicators with each of the three aggregate indicators in turn, are presented in Tables 11, 18, 25 and 32. These tables find that a higher weight, reflected by a higher t -value, is always given to the disaggregate indicator than the aggregate indicator. The disaggregate indicators offer more information about official output growth than the aggregate indicators. However, in general, we cannot always statistically reject $a_2 = 0$; the aggregate indicators do, in general, still offer some information about the official data relative to the disaggregate indicators. Indeed, this is further reflected by the encompassing tests; see Tables 12, 19, 26 and 33. These encompassing tests involve testing: $H_a : a_1 = 1, a_2 = 0$ and $H_b : a_1 = 0, a_2 = 1$. The first hypothesis, a , tests whether the disaggregate indicator encompasses the aggregate indicator, and the second hypothesis, b , tests if the aggregate indicator encompasses the disaggregate indicator. Only in Table 19 when comparing the disaggregate indicators against the Pesaran aggregate indicator is there evidence that the aggregate indicator is encompassed, as evidenced by p -values for H_a greater than 0.25. Otherwise, although the disaggregate indicator offers more information about official output growth than the aggregate indicator, the aggregate indicator still offers some value-added suggesting that it should not be completely disregarded just given a lower weight than the disaggregate indicators.

4.4 Intuition - beginning to understand why the disaggregate indicators work better in-sample

Intuitively we can begin to explain the better performance of the disaggregate approach compared with the aggregate approach as follows. The disaggregate approach combines the responses of the individual firms in a different way to the aggregate approach. The aggregate approach, as we know, simply takes the survey proportions (perhaps weighted by firm size as measured, say, by sales volume) and relates them to the reference series. In contrast the disaggregate approach relates the individual firms to the reference series (via

¹⁵The covariance matrix of the least squares estimator is estimated following Newey and West, 1987. If the forecast errors are not normally distributed, and/or the forecast error processes are time-dependent, heteroscedasticity and autocorrelation robust (HAC) estimation of the covariance matrix is recommended [see Harvey, Leybourne and Newbold, 1998]. In such a situation tests about the regression parameters are incorrectly sized if we use the standard least squares covariance matrix.

the discrete choice model or nonparametrically), quantifies and then aggregates. Thus the disaggregate approach, as is seen from equation (A.2), allows for a degree of heterogeneity among firms not allowed for in the aggregate approach. This different way of aggregating the panel of survey responses may explain the improved performance of the disaggregate approach irrespective of how the firms are weighted in the disaggregate approach.

Turning to this related issue with the disaggregate indicator of how the quantified firm-level series should be aggregated we note that, even in the so-called unweighted case, in a sense the firms are weighted according to how well they have signaled past growth. This is explained in part in 3 (b) above. But let us begin to explain further why, in a sense, the unweighted disaggregate indicator does weight firms according to their track-record. Consider those firms whose responses are poorly correlated with actual output growth. In the extreme case where responses are uncorrelated with output, the inclusion of these reduces the standard deviation of the disaggregate indicator but does not affect its correlation with output growth. This is because for those firms whose responses are orthogonal to output growth, their quantitative series will equal the mean of output growth. Since the correlation of the disaggregate indicator with the reference series is not affected by these firms it is in this sense that the disaggregate indicator “automatically” filters out these poor firms, something not done by the aggregate approach.

It is interesting that this automatic method of attributing a higher weight to the ‘better’ firms delivers a better signal about output growth than explicitly weighting the firms according to their size. It also works better, for Sweden and the UK, than using some empirically based weighting criterion, such as weighting the individual quantified series according to their correlation with output growth itself. This result is perhaps not as surprising as it may first appear. This is because in weighting firms according to their individual correlation (with output growth) we are ignoring the covariances between the firms. This means we are not necessarily taking the “best linear unbiased” combination of the firms’ quantified series, and so we should not necessarily expect the correlation of the weighted indicator to rise. Certainly more work is required on this issue, both empirical and methodological; in particular future work should seek to develop the appropriate theory that would deliver the “optimal” or “best linear unbiased” combination of the firms’ quantified series. In a sense what is required is an extension of the well known “regression-based” method for forecast combination [see Granger and Ramanathan, 1984], that is known to often work well in practice, to the case where the (underlying) data are qualitative and the case where there are more forecasts than time-periods.

5 Comparing the out-of-sample performance of the retrospective aggregate and disaggregate indicators of manufacturing output growth

Having found an improved in-sample fit between the survey responses and official data using the disaggregate rather than aggregate indicators, this section examines whether

the superiority of the disaggregate indicators extends out-of-sample. To evaluate how accurate the survey-based early estimates of output growth would have been out-of-sample we conduct an experiment designed to mimic “real-time” application of the different quantification approaches. We are nevertheless assessing the performance against near-final rather than initial official data.¹⁶

For Germany, Portugal and Sweden the out-of-sample analysis is conducted over the last 3 years; i.e. 36 monthly observations are left for recursive examination. For the U.K. the out-of-sample analysis is conducted over the 8 periods, 1997q4 – 1999q3.¹⁷ The recursive experiments were carried out as follows; we explain the process in detail for the UK.

The aggregate and disaggregate indicators are computed using both survey and official data from 1988q3 to 1997q3, as outlined above, and then these in-sample estimates are used to infer output growth in 1997q4 given knowledge of the survey data in 1997q4, but crucially not the official data on output growth since these are published with a lag.¹⁸ Given that survey data are published ahead of official data this provides an early estimate of output growth. Then data from 1988q3 to 1997q4 are used along with survey data in 1998q1 to infer output growth in 1998q1. This recursive process is carried on until both survey and official data from 1988q3 to 1999q2, plus survey data in 1999q3, are used to infer output growth in 1999q3. Both the aggregate and disaggregate out-of-sample estimates are re-scaled by recursively regressing their in-sample counterparts against the outturn for output growth. In this way no *ex post* information about output growth is used when quantifying the survey data in real-time. As is traditional when evaluating forecasts, the performance of the aggregate and disaggregate indicators is evaluated in terms of their root MSE (RMSE) against the outturn.

The results of this recursive exercise are summarised in Figures 9-12. These figures plot the root MSE of the aggregate indicators (computed using the full-sample) and the disaggregate indicators, both in the unweighted case, against the outturn for manufacturing output growth. The performance of the disaggregate indicators is evaluated as a function of the minimum number of observations considered per firm. For Germany, only

¹⁶We do not have access to real-time data for output growth. Given that real-time (output, GDP) data often provide biased estimates of the final (revised) data, we are not sure that anything general can be inferred about the real-time data from the performance of the indicators against the final data.

¹⁷As indicated above, unfortunately it was not possible to extend the out-of-sample analysis for the UK beyond 1999 since in December 1999 the CBI moved to a new survey processing platform that involved changing the participant identification numbers.

¹⁸For the monthly data (in the cases of Germany, Sweden and Portugal) we assume that when the time t value of the survey data is used to infer output growth in period t , relative to $(t - 3)$, official data on the index of output are available up to period $(t - 1)$. In fact it appears that this information about the official data is only available about ten days after the survey data are published - the survey data are published at the end of the month attributed to it, and are published about ten days ahead of the manufacturing output index for the previous month. Therefore in theory one could obtain an earlier indicator of output growth in period t by using official data only up to period $(t - 2)$. Waiting ten days seems reasonable, however, not just because we should expect the wait to deliver better estimates but because we should expect it would in practice take time to distribute and process the firm-level data ready for computation of the disaggregate indicator.

four cut-off values are considered, given the computational burden of recursively computing the aggregate and disaggregate indicators with close to 4000 firms on average present each month. Nevertheless, this does appear to be sufficient to illustrate the properties of the alternative indicators.

The disaggregate indicators are computed by focusing on those firms present at least a given number of times (as given by the cut-off parameter) in the in-sample period; i.e. firms are not allowed to enter the disaggregate indicators in the out-of-sample period. Results were robust to letting “new” firms enter the sample during the out-of-sample period. Moreover, the parameters in the ordered logit model, used as the basis for the parametric disaggregate indicator, are not updated recursively during the out-of-sample period but just estimated once at the beginning. Experimentation suggested that results with recursive updating were qualitatively similar to those presented here.

Figures 9 and 12 indicate, for Germany and particularly the UK, that both the parametric and nonparametric disaggregate indicators produce more accurate forecasts than the aggregate indicators. Moreover, there is evidence to suggest for the UK that these improvements are statistically significant using small-sample corrected Diebold-Mariano tests; see Harvey *et al.* (1997). This is encouraging and further motivates the use of disaggregate survey based indicators. In the case of Sweden, the parametric disaggregate indicator beats the aggregate indicators for selected values of the cut-off parameter. However, for Portugal, both the parametric and nonparametric disaggregate indicators produce less accurate forecasts than the aggregate indicators.

We should qualify this mixed support for the use of the disaggregate indicators on an out-of-sample basis by noting that out-of-sample analysis, particularly with small-samples, is always sensitive to the period chosen. Indeed in Portugal the out-of-sample period is clearly seen to be characterised by considerable volatility; see Figure 2.

Further evidence on the out-of-sample performance of the aggregate and disaggregate indicators is seen in Table 3. This table focuses on the behaviour of the disaggregate indicators for specific cut-off values. The table again explores whether using some empirically based weighting criterion helps deliver improved forecasts. Table 3 shows that, just like in-sample, it does not; the weighted forecasts are worse than those using equal weights. Table 3 also shows that none of the forecasts beat the unconditional mean over the out-of-sample period; i.e. *s.d.* (namely the standard deviation of output growth during the out-of-sample period) is always lowest. *s.d.* is of course a cheat, since the unconditional mean is only known *ex post*. Encouragingly the survey based indicators are more accurate than a commonly used benchmark time-series model, *DAR*, that is available *ex ante*. *DAR* is a first order autoregressive model in the growth rate of output growth:

$$x_t = x_{t-1} + \tau_t, \tag{2}$$

where τ_t is a mean zero disturbance. It is well known that this model can guard against unforeseen events such as structural breaks; see Clements and Hendry (1999).

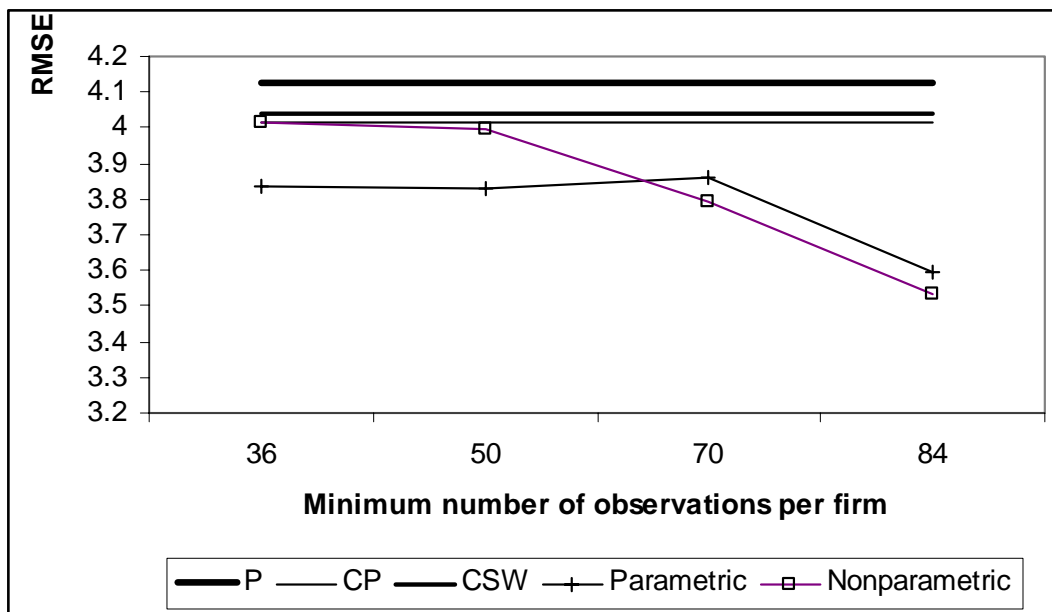


Figure 9: Germany. Out-of-sample performance of the aggregate and disaggregate indicators in the retrospective case. Root MSE of the indicators against manufacturing output growth as a function of the minimum number of observations considered per firm

6 Comparing the in-sample performance of the prospective aggregate and disaggregate indicators of manufacturing output growth

This section considers the in-sample performance of the prospective forecasts of manufacturing output growth. As discussed in Section 2.4, except for the UK, we confine attention to the nonparametric disaggregate indicator. We, of course, continue to compare the informational content of the disaggregate forecasts against their aggregate counterparts. Let us consider the UK first; for more details see Mitchell, Smith and Weale (2004).

Table 4 compares the in-sample predictive power of parametric and nonparametric disaggregate indicators against two of the aggregate approaches (Carlson-Parkin and Pesaran) in the UK, focusing on a cut-off value of twenty. Both unweighted and size-weighted results are considered. Table 4 does make clear that the parametric and nonparametric disaggregate forecasts, \hat{x}_t^D and \hat{x}_t^{ND} , explain more of the variation in output growth one quarter ahead than the aggregate indicators. The disaggregate indicators provide more accurate leading indicators of output growth than traditional aggregate indicators.

Figures 13-16 summarise the performance of the nonparametric disaggregate forecast for a series of cut-off values for Germany, Portugal, Sweden and the UK. These figures clearly show that, as in Table 4, an improved signal about future movements in output growth can be derived from the disaggregate indicators: the correlation of the disaggre-

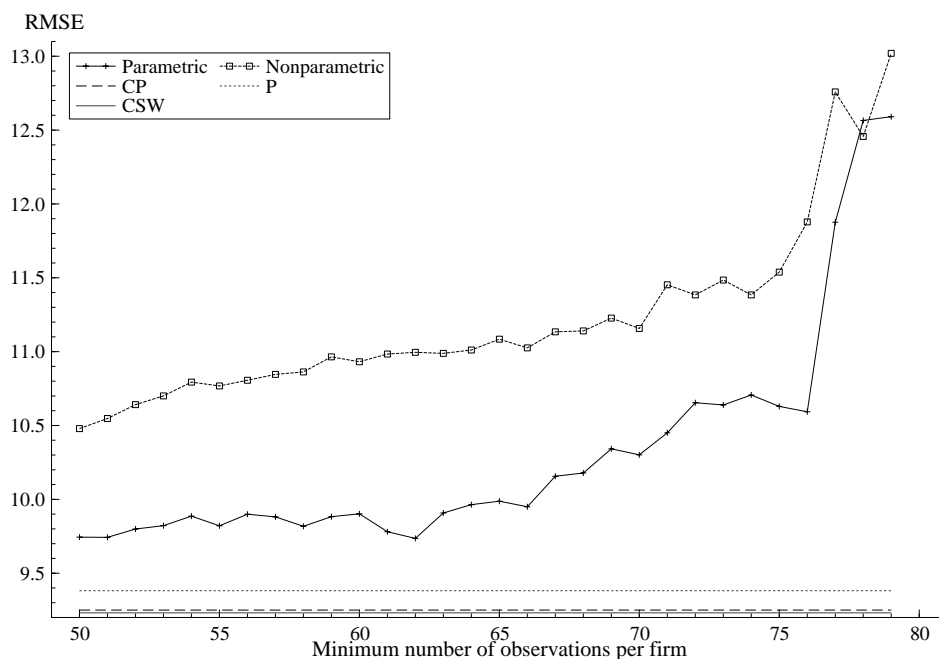


Figure 10: Portugal. Out-of-sample performance of the aggregate and disaggregate indicators in the retrospective case. Root MSE of the indicators against manufacturing output growth as a function of the minimum number of observations considered per firm

gate indicator with the subsequent outturn for output growth is higher when using the disaggregate forecast than the aggregate forecasts for all cut-off values.

7 Comparing the out-of-sample performance of the prospective aggregate and disaggregate indicators of manufacturing output growth

This section examines whether the superiority of the disaggregate indicators in the prospective case extends out-of-sample. As in Section 5 we conduct a series of recursive simulations in the out-of-sample period. However, one difference with the out-of-sample analysis conducted in the retrospective case, is that in the prospective case we must acknowledge the greater lag in the availability of official data on output growth; see Section C.4. Let us explain this for the UK.

For the UK the out-of-sample analysis is again conducted using the prospective survey responses over the 8 periods, 1997q4 – 1999q3, so that forecasts for output growth are

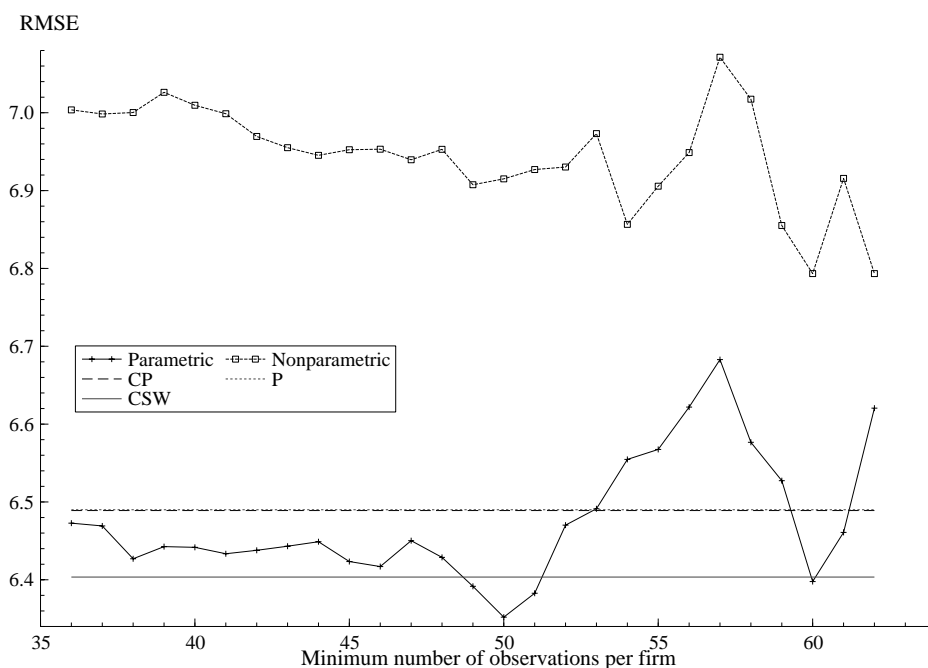


Figure 11: Sweden. Out-of-sample performance of the aggregate and disaggregate indicators in the retrospective case. Root MSE of the indicators against manufacturing output growth as a function of the minimum number of observations considered per firm

obtained for 1998 $q1$ – 1999 $q4$. Just as with the in-sample analysis conducted above, on an out-of-sample basis we relate the survey data published at quarter t (but assumed to refer to $(t + 1)$) to official data for quarter $(t + 1)$. However, out-of-sample we need to reflect the fact that the official data for output growth are published with a lag. Indeed, as seen above in Sections 4 and 5 we have already found that the retrospective survey responses (published at time t and referring to t) can be exploited to obtain useful ‘early’ estimates of these official data, given that the survey data are published ahead of the official data; for more details see Mitchell, Smith and Weale (2002a).

The analysis is performed by conducting the following recursive experiments. When using the prospective survey responses published in 1997 $q4$ to forecast output growth in 1998 $q1$ since the official data for 1997 $q4$ are assumed not yet published, the in-sample estimates, used as the basis for the out-of-sample forecasts, relate the prospective survey data published in 1988 $q3$ to 1997 $q2$ to official data for 1988 $q4$ to 1997 $q3$. Section C.4 details how the parametric disaggregate indicator is made operational out-of-sample. A similar delay is used in the application of the aggregate methods. Then we forecast output growth for 1998 $q2$ using the prospective data published in 1998 $q1$, given the in-sample estimates based on relating the prospective survey data from 1988 $q3$ to 1997 $q3$

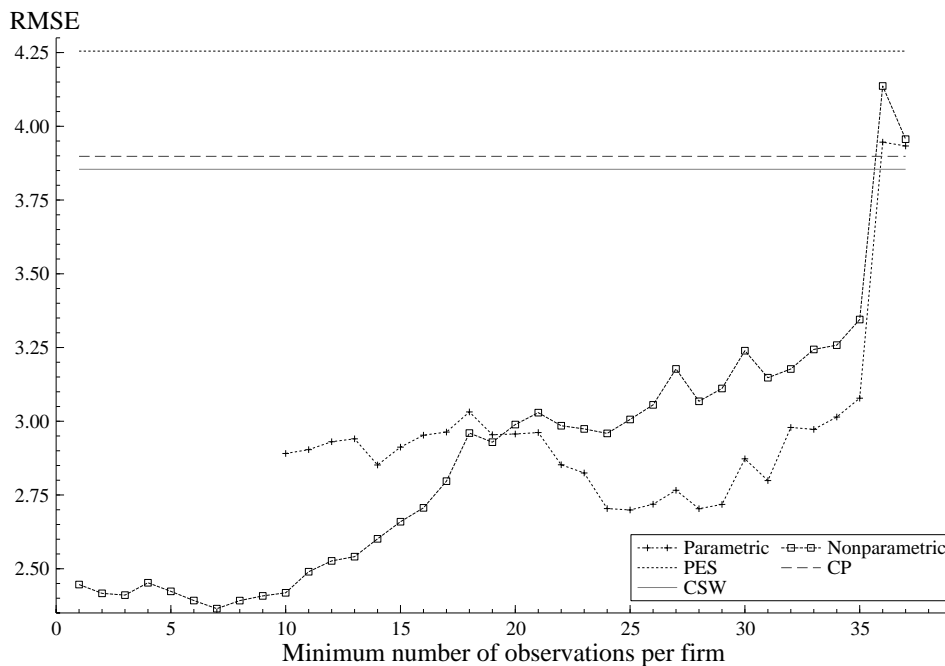


Figure 12: U.K. Out-of-sample performance of the aggregate and disaggregate indicators in the retrospective case. Root MSE of the indicators against manufacturing output growth as a function of the minimum number of observations considered per firm

to official data from 1988 $q4$ to 1997 $q4$. This recursive process is carried on until survey data published in 1999 $q3$ are used to forecast output growth in 1999 $q4$, given in-sample estimates based on relating the survey data published in 1988 $q3$ to 1999 $q1$ to official data for 1988 $q4$ to 1999 $q2$. Both the aggregate and disaggregate out-of-sample estimates are re-scaled by recursively regressing their in-sample counterparts against the outturn for output growth following (C.21) and (C.22); we denote these forecasts \hat{x}_t^D and \hat{x}_t^{ND} . In this way no *ex post* information about output growth is used when quantifying the survey data in “real-time”.

For the monthly data (in the cases of Germany, Portugal and Sweden) we assume that at month t when using the prospective survey data published in month t to forecast growth relative to month $(t + 3)$ official data on output growth are available up to month $(t - 1)$.

Let us again consider the UK first, focusing on the cut-off value of twenty, as for the UK we have considered both the parametric and nonparametric disaggregate forecasts; see Table 5. The first conclusion from Table 5 is that the nonparametric disaggregate indicator produces more accurate forecasts than the aggregate indicators. Moreover, there is evidence to suggest that these improvements are statistically significant using small-

Table 3: Out-of-sample performance as measured by RMSE of the aggregate and disaggregate indicators in the retrospective case subject to alternative means of weighting the firm-level series when computing the disaggregate indicator

RMSE	Germany	Portugal	Sweden	UK
	cut-off=84	cut-off=50	cut-off=50	cut-off=20
CP	4.95	9.25	6.49	3.90
P	4.12	9.38	6.49	4.25
CSW	4.04	9.23	6.40	3.85
BAL	4.01	9.25	6.47	3.90
N	3.60	9.74	6.35	2.83
NP	3.53	10.48	6.92	2.74
N: R2	3.88	9.77	6.34	3.40
N: poly	3.94	9.65	6.37	3.96
NP R2	3.56	9.93	6.81	2.91
NP poly	3.61	9.80	6.52	3.38
N: RMSE	4.53	10.70	6.94	3.96
N:cor	3.86	10.31	6.58	3.63
NP: RMSE	3.62	11.31	7.48	3.35
NP: cor	-	11.63	-	-
DAR	4.26	13.18	7.85	2.93
s.d.	4.04	9.01	5.98	2.63

Notes: N denotes the parametric disaggregate indicator, denoted D_t in Appendix A; NP denotes the nonparametric disaggregate indicator, denoted ND_t in Appendix A; R2 denotes that the disaggregate indicator has been derived by weighting firm's quantified series according to the pseudo R^2 of the ordered logit model; poly denotes that the disaggregate indicator has been derived by weighting firm's quantified series according to the polyserial correlation between the categorical responses in the survey and the official data; RMSE and cor denote that the disaggregate indicator has been derived by weighting firm's quantified series according to its RMSE or correlation against the official data; DAR is the auto-regressive model and s.d. is the standard deviation of output growth during the out-of-sample period

sample corrected Diebold-Mariano tests; see Harvey *et al.* (1997). This is encouraging and further motivates the use of disaggregate survey based indicators. The nonparametric disaggregate survey based forecasts also beat those of DAR , the benchmark time-series model. This is particularly encouraging given that these time-series forecasts are in fact using more information than would be available in practice. However, the parametric disaggregate indicator does not deliver as accurate forecasts as its nonparametric cousin. Its forecasts are only marginally more accurate than the aggregate indicators, and only then when new firms are allowed to enter the sample during the out-of-sample period. More work is required to consider whether the improved in-sample performance of the parametric disaggregate indicator, relative to the aggregate indicator, can be translated into better performance out-of-sample. The manner in which this indicator is re-scaled should be central to this.

Turning to the performance of the nonparametric disaggregate forecasts in the other

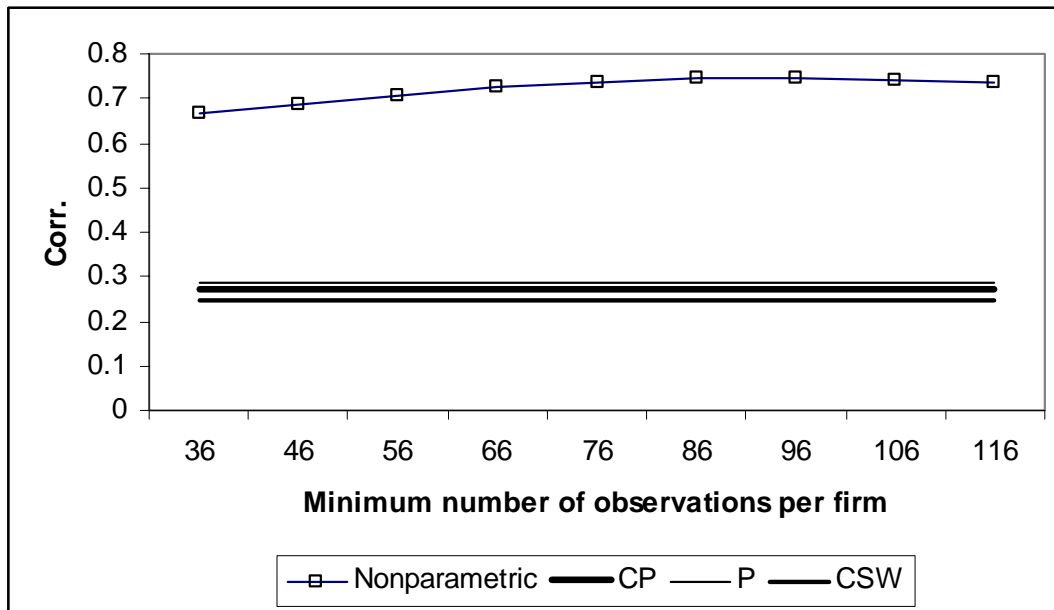


Figure 13: Germany. In-sample performance of the aggregate and disaggregate indicators in the prospective case. Correlation of the forecasts against manufacturing output growth as a function of the minimum number of observations considered per firm

countries Figures 17-20 compare the predictive power of the disaggregate forecast against those of the aggregate forecasts for a range of cut-off values. They clearly show that, except for Sweden, the disaggregate forecasts deliver more accurate forecasts than their aggregate counterparts, in the sense of a lower RMSE. There is some deterioration in performance of the disaggregate forecasts for Portugal for high values of the cut-off parameter; but this is perhaps not surprising as fewer and fewer firms are left in the sample. However, although there are gains in the use of the disaggregate forecast, except for perhaps the UK, they are not statistically significant. More work is required to consider whether the improved in-sample performance of the disaggregate indicators, relative to the aggregate indicator, can be translated into better performance out-of-sample. Both the manner in which this indicator is re-scaled, and the way in which the firm-level quantified series are weighted, should be central to this.

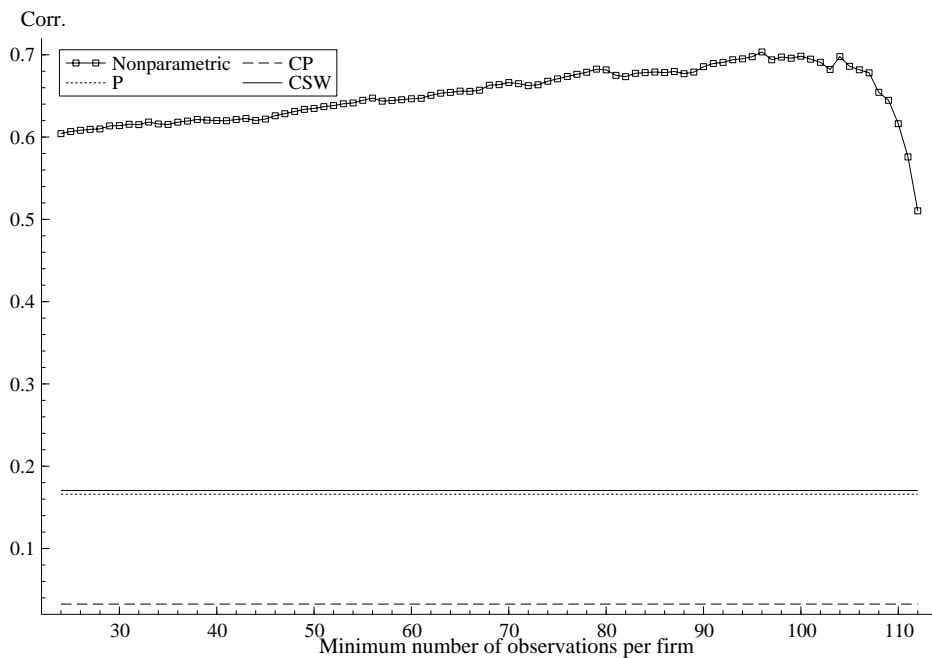


Figure 14: Portugal. In-sample performance of the aggregate and disaggregate indicators in the prospective case. Correlation of the forecasts against manufacturing output growth as a function of the minimum number of observations considered per firm

Table 4: Aggregate and Disaggregate Forecast Performance in the Prospective Case: Detailed In-sample Results for the UK: 1988q4 – 1997q4

		Mean	Stand. Dev.	Corr.
Outturn for Manuf. Output Growth		0.807	3.951	
CP	unweighted	0.807	1.333	0.627
	size-weighted	0.807	1.493	0.625
P	unweighted	0.807	2.628	0.665
	size-weighted	0.807	2.583	0.654
\hat{x}_t^D	unweighted	0.807	3.580	0.906
	size-weighted	0.807	3.503	0.887
\hat{x}_t^{ND}	unweighted	0.807	3.559	0.900
	size-weighted	0.807	3.555	0.899

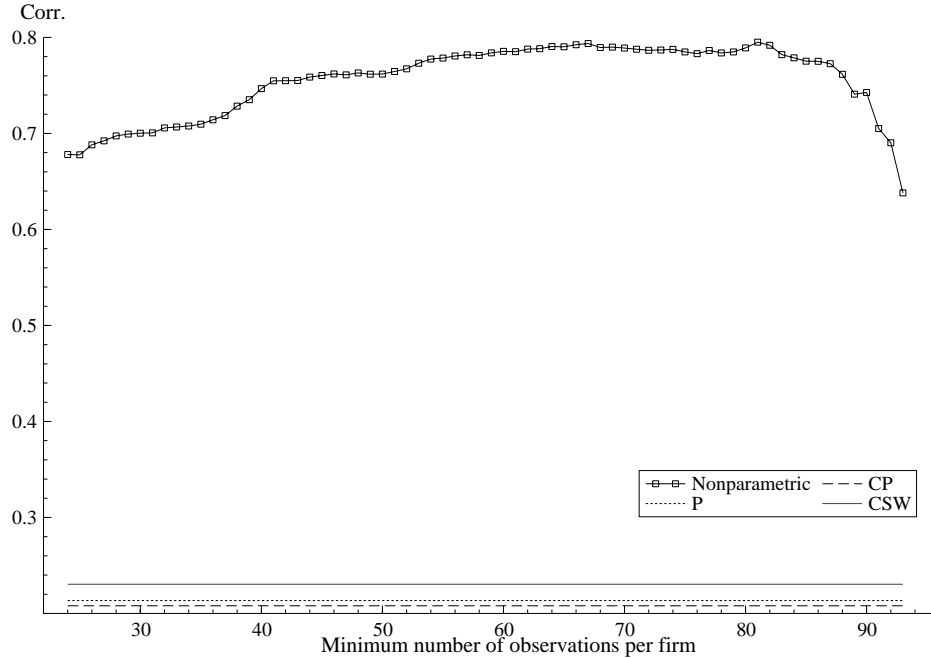


Figure 15: Sweden. In-sample performance of the aggregate and disaggregate indicators in the prospective case. Correlation of the forecasts against manufacturing output growth as a function of the minimum number of observations considered per firm

Table 5: Aggregate and Disaggregate Forecast Performance in the Prospective Case: Out-of-sample results for the UK, 1998q1 – 1999q4

	Root MSE
CP	3.763
P	3.906
\hat{x}_t^D	3.732
\hat{x}_t^{ND}	2.117
DAR	3.065

Notes. The results for the disaggregate forecasts presented here allow “new” firms to enter the sample during the out-of-sample period. If we restrict attention in the out-of-sample analysis to those 693 firms present in the in-sample period similar results are obtained; the parametric and nonparametric disaggregate forecasts now have root MSE, respectively, of 3.850 and 2.298, rather than 3.732 and 2.117.

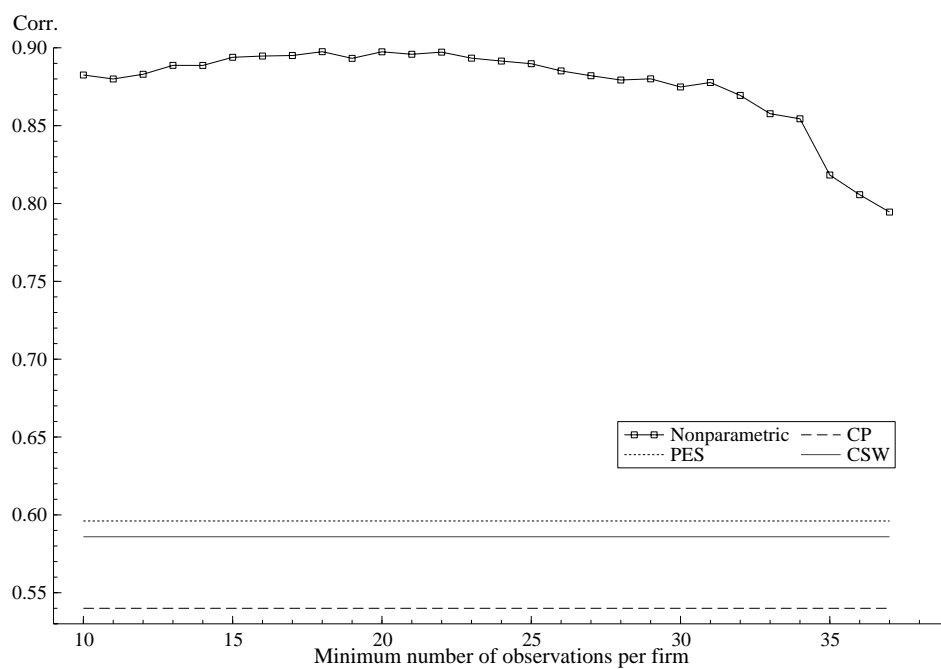


Figure 16: UK. In-sample performance of the aggregate and disaggregate indicators in the prospective case. Correlation of the forecasts against manufacturing output growth as a function of the minimum number of observations considered per firm

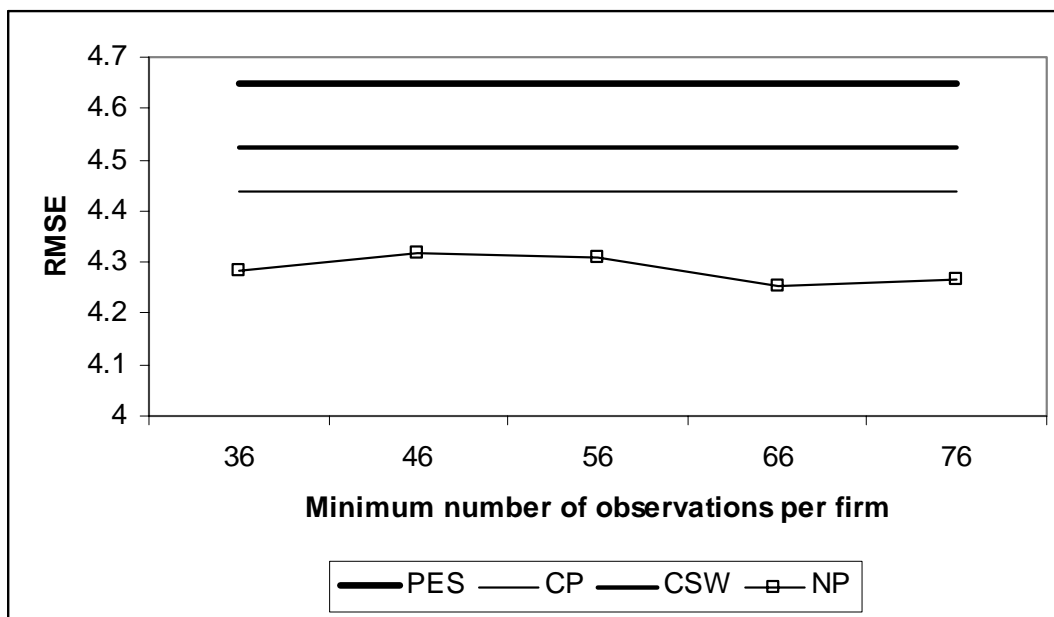


Figure 17: Germany. Out-of-sample performance of the aggregate and disaggregate indicators in the prospective case. Root MSE of the forecasts against manufacturing output growth as a function of the minimum number of observations considered per firm

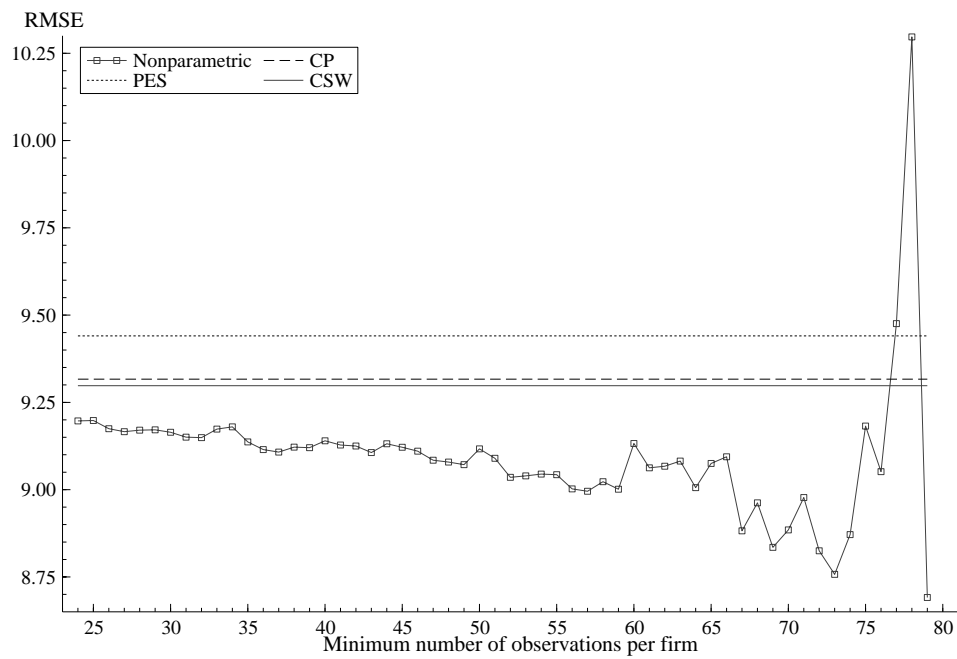


Figure 18: Portugal. Out-of-sample performance of the aggregate and disaggregate indicators in the prospective case. Root MSE of the forecasts against manufacturing output growth as a function of the minimum number of observations considered per firm

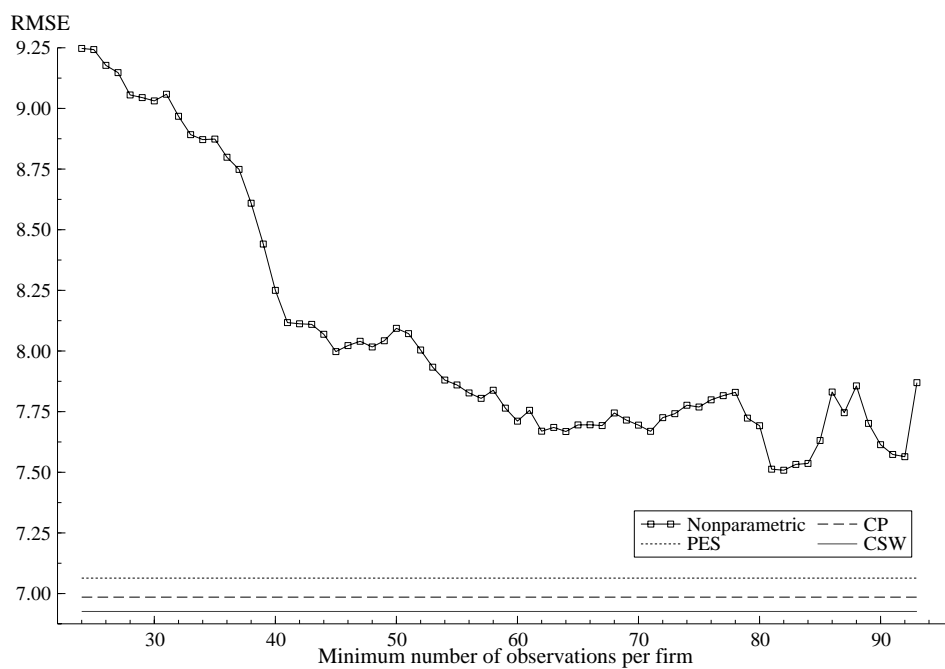


Figure 19: Sweden. Out-of-sample performance of the aggregate and disaggregate indicators in the prospective case. Root MSE of the forecasts against manufacturing output growth as a function of the minimum number of observations considered per firm

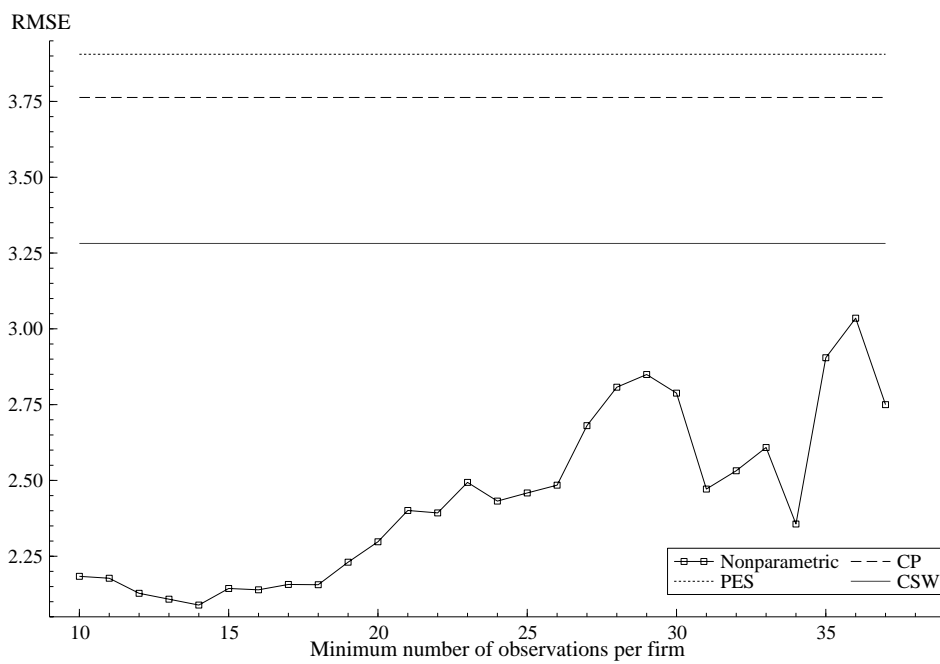


Figure 20: UK. Out-of-sample performance of the aggregate and disaggregate indicators in the prospective case. Root MSE of the forecasts against manufacturing output growth as a function of the minimum number of observations considered per firm

8 References

References

- [1] Albert, J.H. and S. Chib (1993), “Bayesian analysis of binary and polychotomous response data”, *Journal of the American Statistical Association*, 88, 669-679.
- [2] Amemiya, T. (1981), “Qualitative response models: a survey”, *Journal of Economic Literature*, 19, 1483-1536.
- [3] Amemiya, T. (1985), *Advanced Econometrics*, Basil Blackwell.
- [4] Anderson, O. (1952), “The business test of the IFO-Institute for Economic Research, Munich, and its theoretical model”, *Review of the International Statistical Institute*, 20, 1-17.
- [5] Batchelor, R. (1981), “Aggregate expectation under the stable laws”, *Journal of Econometrics*, 16, 199-210.
- [6] Batchelor, R. (1982), “Expectations, output and inflation”, *European Economic Review*, 17, 1-25.
- [7] Branch, W.A. (2004), “The theory of rationally heterogeneous expectations: evidence from survey data on inflation expectations”, *Economic Journal*, 114, 592-621.
- [8] Britton, E., J. Cutler and A. Wardlow (1999), “The Bank’s use of survey data”, *Bank of England Quarterly Bulletin*, May, 177-182.
- [9] Carlson, J.A. and M. Parkin (1975), “Inflation expectations”, *Economica*, 42, 123-138.
- [10] Chesher, A. and M. Irish (1987), “Residual analysis in the grouped and censored normal linear model”, *Journal of Econometrics*, 34, 33-61.
- [11] Clements, M.P. and D.F. Hendry (1999). *Forecasting Non-Stationary Time Series*. Cambridge, MIT Press.
- [12] Coenders, G., A. Satorra and W.E. Saris (1997), “Alternative approaches to structural modelling of ordinal data: A Monte-Carlo study”, *Structural Equation Modeling*, 4(4), 261-282.
- [13] Cunningham, A.W.F. (1997), “Quantifying survey data”, *Bank of England Quarterly Bulletin*, August, 292-300.
- [14] Cunningham, A.W.F., R.J. Smith and M.R. Weale (1998), “Measurement errors and data estimation: the quantification of survey data”. In *Applied Economics and Public Policy* (eds. Begg, I. and S.G.B. Henry), Cambridge University Press.

- [15] Granger, C.W.J. and R. Ramanathan (1984), “Improved methods of combining forecasts”, *Journal of Forecasting*, 3, 197-204.
- [16] Harvey, D. I. and S. J. Leybourne and P. Newbold (1997), “Testing the Equality of Prediction Mean Square Errors”, *International Journal of Forecasting*, 13, 273–281.
- [17] Harvey, D.I., S.J. Leybourne and P. Newbold (1998), “Tests for forecast encompassing”, *Journal of Business and Economic Statistics*, 16, 254-259.
- [18] Horvath, B., M. Nerlove and D. Willson (1992). “A reinterpretation of direct tests of forecast rationality using business survey data”. In *Business Cycle Analysis by Means of Economic Surveys: Part I* (eds. Oppenländer, K.H. and Poser, G.). Avebury, Aldershot.
- [19] Hsiao, C. (2003), *Analysis of Panel Data* (Second edition), Cambridge University Press.
- [20] Keane, M.P. (1994), “A Computationally Practical Simulation Estimator for Panel Data”, *Econometrica*, 62, 95-116.
- [21] Lee, K.C. (1994), “Formation of price and cost inflation expectations in British manufacturing industries: a multisectoral analysis”, *Economic Journal*, 104, 372-386.
- [22] Machin, S.J. and M.B. Stewart (1990), “Unions and the financial performance of British private sector establishments”, *Journal of Applied Econometrics*, 5, 327-350.
- [23] McIntosh, J., F. Schiantarelli and W. Low (1989), “A qualitative response analysis of UK firms’ employment and output decisions”, *Journal of Applied Econometrics*, 4, 251-264.
- [24] Mitchell, J. (2002), “The use of non-normal distributions in quantifying qualitative survey data on expectations”, *Economics Letters*, 76, 101-107.
- [25] Mitchell, J., R.J. Smith and M.R. Weale (2002a), “Aggregate versus disaggregate survey-based indicators of economic activity”, Available as NIESR Discussion Paper 194. Currently under revision for publication.
- [26] Mitchell, J., R.J. Smith and M.R. Weale (2002b), “Quantification of qualitative firm-level survey data”, *Economic Journal*, 112, 117-135.
- [27] Mitchell, J., R.J. Smith and M.R. Weale (2004), “Forecasting manufacturing output growth using firm-level survey data”, mimeo, NIESR.
- [28] Murphy, A. (1996), “Simple LM tests of mis-specification for ordered logit models”, *Economics Letters*, 52, 137-141.
- [29] Nerlove, M. (1983). “Expectations, plans, and realizations in theory and practice”. *Econometrica*, vol. 51, 1251-1279.

- [30] Newey, W.K. (1987), "Efficient estimation of limited dependent variable models with endogenous explanatory variables", *Journal of Econometrics*, 36, 231-250.
- [31] Newey, W.K. and K.D. West (1987), "A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, 55, 703-708.
- [32] Olsson, U., F. Drasgow and N.J. Dorans (1982), "The polyserial correlation coefficient", *Psychometrika*, 47, 337-347.
- [33] Pesaran, M.H. (1984), "Expectations formation and macroeconomic modelling". In *Contemporary Macroeconomic Modelling* (eds. Malgrange, P. and P. Muet), Blackwell.
- [34] Pesaran, M.H. (1987), "The limits to rational expectations", Basil Blackwell.
- [35] Smith, J. and M. McAleer (1995), "Alternative procedures for converting qualitative response data to quantitative expectations: an application to Australian manufacturing", *Journal of Applied Econometrics*, 10, 165-185.
- [36] Smith, R.J. and R.W. Blundell (1986), "An exogeneity test for a simultaneous equation Tobit model with an application to labor supply", *Econometrica*, 54, 679-685.
- [37] Souleles, N.S. (2004), "Expectations, heterogeneous forecast errors, and consumption: Micro evidence from the Michigan consumer sentiment surveys", *Journal of Money, Credit and Banking*, 36, 39-72.
- [38] Stone, J.R.N., D.A. Champernowne and J.E. Meade (1942), "The precision of national income estimates", *Review of Economic Studies*, 9, 111-25.
- [39] Theil, H. (1952), "On the time shape of economic microvariables and the Munich business test", *Revue de l'Institute Internationale de Statistique*, 20, 105-120.
- [40] Weale, M.R. (2004), "Box B - Business surveys and manufacturing output", *National Institute Economic Review*, 188, 48-49.
- [41] Wren-Lewis, S. (1985), "The quantification of survey data on expectations", *National Institute Economic Review*, 113, 39-49.

A Disaggregate indicator: technical details about firm-level quantification

This appendix provides some of the technical details behind the disaggregate indicator proposed by Mitchell, Smith and Weale (2002a) for examination of the retrospective survey responses. For the convenience of the reader it is re-produced from Mitchell, Smith and Weale (2002a).

A.1 Ordered Discrete Choice Models

Consider a survey that asks a sample of N_t manufacturing firms at time t whether their output growth, for example, has risen, not changed or fallen relative to the previous period. Crucially the number of firms in the sample is allowed to vary across t . The categorical responses in the survey are assumed to relate to observed official data for economy-wide manufacturing output growth x_t in the following manner. Let the actual output growth of firm i at time t , y_{it} , ($i = 1, \dots, N_t$), depend on x_t according to the conditional linear model

$$y_{it} = x_t + \eta_{it} + \varepsilon_{it}, \quad (\text{A.1})$$

($t = 1, \dots, T$), where η_{it} is the difference between y_{it} and x_t anticipated by firm i , reflecting information private to firm i at time t that is not observed by the econometrician. This information may reflect firm or industry level influences. The random variable ε_{it} captures the component of firm-specific output growth y_{it} unanticipated by both firm i and the econometrician at time t . That is, $E(y_{it}|\Omega_t^i) = x_{it} = x_t + \eta_{it}$, where Ω_t^i comprises information available to firm i at time t and includes x_t , and $E(y_{it}) = x_t$. We may re-express (A.1) as

$$y_{it} = \alpha_i + \beta_i x_t + \varepsilon_{it}, \quad (\text{A.2})$$

where α_i and β_i are firm-specific time-invariant coefficients that can be expressed in terms of (A.1) by defining $\eta_{it} = \alpha_i + (\beta_i - 1)x_t$, ($i = 1, \dots, N_t$, $t = 1, \dots, T$). It follows that $E(y_{it}|\Omega_t^i) = x_t + \eta_{it} = \alpha_i + \beta_i x_t$ so $E(\varepsilon_{it}|\Omega_t^i) = 0$ and ε_{it} is uncorrelated with x_t .¹⁹ Therefore, by assumption x_t is weakly exogenous. Nevertheless, empirical validation for this assumption, and more generally for the specification (A.2), is sought below by subjecting (A.2) to a range of mis-specification tests; see Section A.2 below. Indeed the model (A.2) can be straightforwardly augmented to accommodate x_t being endogenous

¹⁹Furthermore, $E(y_{it}|\Omega_t) = \alpha + \beta x_t$ where $E(\alpha_i|\Omega_t) = \alpha$, $E(\beta_i|\Omega_t) = \beta$ and Ω_t comprises information available to all firms at time t and includes x_t . Let z_{it} denote (the level of) output of firm i at time t . From (A.2), $\sum_{i=1}^{N_t} \Delta z_{it} = \sum_{i=1}^{N_t} z_{it-1} \alpha_i + \sum_{i=1}^{N_t} z_{it-1} \beta_i x_t + \sum_{i=1}^{N_t} z_{it-1} \varepsilon_{it}$, after cross-multiplication and summation over $i = 1, \dots, N_t$, where Δ is the first difference operator. For coherency we require $\sum_{i=1}^{N_t} \Delta z_{it} / \sum_{i=1}^{N_t} z_{it-1} \xrightarrow{p} x_t$, $\sum_{i=1}^{N_t} z_{it-1} \alpha_i / \sum_{i=1}^{N_t} z_{it-1} \xrightarrow{p} 0$, $\sum_{i=1}^{N_t} z_{it-1} \beta_i / \sum_{i=1}^{N_t} z_{it-1} \xrightarrow{p} 1$ and $\sum_{i=1}^{N_t} z_{it-1} \varepsilon_{it} / \sum_{i=1}^{N_t} z_{it-1} \xrightarrow{p} 0$ ($N_t \rightarrow \infty$).

as well as driven by a dynamic process. In the following analysis it is further assumed that output growth x_t is a stationary variable.

Actual growth y_{it} of firm i at time t is unobserved but the survey contains data corresponding to whether output growth has risen, not changed or fallen relative to the previous period. To account for the ordinal nature of the responses, we use ordered discrete choice models [see Amemiya (1985), Ch.9] based on the latent regression (A.2). Define the indicator variables

$$y_{it}^j = 1 \text{ if } \mu_{(j-1)i} < y_{it} \leq \mu_{ji} \text{ and } 0 \text{ otherwise, } (j = 1, 2, 3), \quad (\text{A.3})$$

corresponding to “down”, “same” and “up”, respectively, where $\mu_{0i} = -\infty$, μ_{1i} , μ_{2i} and $\mu_{3i} = \infty$ are firm-specific threshold parameters.²⁰ We assume that the error terms ε_{it} , ($t = 1, \dots, T$), are logistic with common cumulative distribution function (c.d.f.) $F(z) = [1 + \exp(-z)]^{-1}$, $-\infty < z < \infty$, ($i = 1, \dots, N_t$).²¹ The probabilistic foundation for the observation rule (A.3) is given by the conditional probability $P_{jit} \equiv P_i(j|x_t, i)$ of observing the categorical response $y_{it}^j = 1$ for choice j at time t given the value of x_t and firm i

$$P_{jit} = F(\mu_{ji} - \alpha_i - \beta_i x_t) - F(\mu_{(j-1)i} - \alpha_i - \beta_i x_t), (j = 1, 2, 3). \quad (\text{A.4})$$

Assuming the errors ε_{it} are independently and identically distributed over time, the likelihood function for firm i is

$$L_i = \prod_{t=1}^T P_{1it}^{y_{it}^1} P_{2it}^{y_{it}^2} P_{3it}^{y_{it}^3}. \quad (\text{A.5})$$

Under the above assumptions, maximisation of (A.5) yields consistent estimates ($T \rightarrow \infty$) of α_i , β_i and μ_{ji} denoted by $\hat{\alpha}_i$, $\hat{\beta}_i$ and $\hat{\mu}_{ji}$ respectively.²² Alternatively rather than estimating *via* maximum likelihood, Bayesian methods such as Markov Chain Monte Carlo could be employed; see Albert and Chib (1993).

A.2 Specification tests

The simple model (A.2) can be interpreted as a restricted case of a more general model that allows for dynamic dependence in the official data; it is important to test empirically the implied restrictions given that macroeconomic data are widely accepted to exhibit dependence. The general model, a two equation simultaneous model, consists of (A.2) plus an auto-regressive process governing the determination of x_t . Without loss of generality let us assume that x_t , measured relative to its mean, follows a first-order auto-regressive

²⁰Discrete choice models are only identified up to scale; including the intercept α_i in (A.2) necessitates setting, for example, the first threshold parameter μ_{1i} to zero to achieve identification.

²¹The logistic distribution is similar in shape to the normal but has slightly heavier tails. The logistic distribution is convenient since it offers a closed form distribution function.

²²As the parameters α_i , β_i and μ_{2i} are only identified up to scale, the decision probabilities (A.4) are invariant to multiplying (A.2) by an arbitrary constant.

process, AR(1),²³

$$x_t = \lambda x_{t-1} + u_t, \quad (\text{A.6})$$

($t = 1, \dots, T$), where $|\lambda| < 1$ to ensure stationarity of output growth and u_t is an *i.i.d.* mean zero disturbance. ε_{it} and u_t follow a bivariate distribution and assume that conditionally on u_t we can write their dependence in the form

$$\varepsilon_{it} = \rho_i u_t + \eta_{it}, \quad (\text{A.7})$$

where ρ_i ($i = 1, \dots, N_t$) is a firm-specific parameter and η_{it} is an *i.i.d.* mean zero logistic disturbance distributed independently of u_t . Substitution in (A.2) generates the conditional model

$$y_{it} = \alpha_i + \beta_i x_t + \rho_i u_t + \eta_{it}. \quad (\text{A.8})$$

A test of $\rho_i = 0$ ($i = 1, \dots, N_t$) amounts to a specification test for (A.2). It is a joint test in the sense that $\rho_i = 0$ implies not just that there are no dynamics in x_t but that x_t is (weakly) exogenous in (A.2). Along the lines of Smith and Blundell (1986) and Newey (1987) we consider a two-step test for $\rho_i = 0$. First, estimate (A.6) by Ordinary Least Squares. This yields a consistent estimate, $\hat{\lambda}$, for λ (as $T \rightarrow \infty$). Secondly, estimate the parameters in (A.8), i.e. α_i , β_i , μ_{ji} and ρ_i , by ordered logit substituting \hat{u}_t for u_t , where $\hat{u}_t = x_t - \hat{\lambda}x_{t-1}$. Failure to reject $\rho_i = 0$ supports use of (A.2) alone, while rejection implies that the official data should be inferred (see Section A.3 below) using the conditional model (A.8). To mitigate the effects of an inflated Type 1 error when testing $\rho_i = 0$ across i ($i = 1, \dots, N_t$) a Bonferroni type correction is used. Additional empirical validation for (A.2) can be sought by subjecting (A.2) to a range of score or Lagrange Multiplier tests of mis-specification (omitted variables, heteroscedasticity, incorrect functional form and asymmetry) appropriate for the ordered logit model; see Chesher and Irish (1987), Machin and Stewart (1990) and Murphy (1996).

While it is important if undertaking structural inference using the estimated logit models to ensure the models adequately explain the data, it is well known that there is no reason to expect good in-sample fit to translate into good forecasts. In a forecasting context it is important to undertake simulated out-of-sample experiments to assess the forecasting performance of the selected models against benchmark forecasts.

A.3 Inferring the Official Data

Given an ordered logit model for each firm i , an estimator for x_t may be inferred from the survey data. As survey data are usually published ahead of the official data, this provides an early quantitative estimate of x_t . Since they are not subject to revision they must be assessed against near-final official data.

Let j_{it} , ($j_{it} = 1, 2, 3$), denote the survey response of firm i at time t , where 1, 2 and 3 correspond to “down”, “same” and “up”, respectively. Our initial interest centres on

²³Additional auto-regressive terms can be included if necessary to render u_t serially uncorrelated. In fact, empirically while an AR(1) was chosen for the UK, for Germany, Portugal and Sweden, where data are monthly rather than quarterly, an AR(12) was used.

the conditional density $f(x_t|j, i)$ for observing x_t given the survey response j for firm i . Let $f(x_t)$ denote the time-invariant density function of x_t . Therefore, the conditional probability of observing response j for firm i is $P(j|i) = \int_{-\infty}^{\infty} P(j|x_t, i)f(x_t)dx_t$. Bayes' Theorem states that

$$f(x_t|j, i) = \frac{P(j|x_t, i)f(x_t)}{P(j|i)}. \quad (\text{A.9})$$

For firm i , the Bayes estimator (under squared error loss) for x_t given j is the mean of the posterior density $f(x_t|j, i)$:

$$E(x_t|j, i) = \int_{-\infty}^{\infty} x_t f(x_t|j, i) dx_t, \quad (\text{A.10})$$

which takes one of three values depending on the observed sample response j_{it} of firm i at time t . Given $f(x_t)$, all of the above integrals may be calculated by numerical evaluation.

Estimators $\hat{P}(j|x_t, i)$ for $P(j|x_t, i)$ and, thus, $\hat{P}(j|i)$ for $P(j|i)$ are given by substitution of the estimators $\hat{\alpha}_i$, $\hat{\beta}_i$ and $\hat{\mu}_{ji}$, ($j = 0, \dots, 3$), in (A.4). Hence, a feasible Bayes estimator $\hat{E}(x_t|j, i)$ may be obtained from (A.10) by numerical evaluation.

To create a disaggregate indicator D_t of economic activity at time t , from the law of iterated expectations the conditional expectation of x_t given all firms' survey responses j_{it} , ($i = 1, \dots, N_t$),

$$E(x_t|\{j_{it}\}_{i=1}^{N_t}) = \sum_{i=1}^{N_t} H_{it} E(x_t|j_{it}, i), \quad (\text{A.11})$$

where H_{it} is the exogenous sample probability of observing firm i at time t . Hence, assuming firms are independent, we define the parametric indicator

$$D_t = \sum_{i=1}^{N_t} w_{it} \hat{E}(x_t|j_{it}, i), \quad (\text{A.12})$$

where $w_{it} > 0$ is the weight given to firm i at time t and $\sum_{i=1}^{N_t} w_{it} = 1$. If firms constitute a random sample, then equal weights are appropriate since all firms are equally likely in the sample. However, if firms are drawn according to some stratified sampling process, then the weights w_{it} should reflect stratum weights; for example, if strata are defined by firm size, then firms should be size-weighted.

An alternative non-parametric disaggregate indicator ND_t for the conditional expectation $E(x_t|\{j_{it}\}_{i=1}^{N_t})$ which avoids the assumption of a parametric structure for $f(x_t|j, i)$ via (A.9) may be based on the conditional empirical distribution function. Define the indicator function $1(x_t \leq x, j_{it} = j|i) = 1$ if $x_t \leq x$ and $j_{it} = j$ and 0 otherwise, ($j = 1, 2, 3$). Let $T_i^j = \sum_{s=1}^T y_{is}^j$ which is the number of times firm i gives response j in the survey; hence, T_i^j/T is the sample proportion of responses j for firm i , ($j = 1, 2, 3$). The conditional empirical distribution function of x_t given response j for firm i is given by $\hat{F}(x|j, i) = \sum_{t=1}^T 1(x_t \leq x, j_{it} = j|i)/T_i^j$, ($j = 1, 2, 3$), which assigns equal weight to each sample value. As $T \rightarrow \infty$ and, thus, $T_i^j \rightarrow \infty$, $T_i^j/T \xrightarrow{P} P(j|i)$ and $\sum_{t=1}^T 1(x_t \leq x, j_{it} = j|i)/T \xrightarrow{P} F(x, j|i)$ if, given firm i , x_t and j_{it} may be regarded as stationary random variables with joint conditional c.d.f. $F(x, j|i)$. Hence,

$\hat{F}(x|j, i) \xrightarrow{p} F(x|j, i) = F(x, j|i)/P(j|i)$, the conditional c.d.f. of x_t given response j and firm i . Therefore, the mean of $\hat{F}(x|j, i)$, $\sum_{s=1}^T y_{is}^j x_s / T_i^j$, is a consistent estimator for $E(x_t|j, i)$. A nonparametric disaggregate (ND_t) indicator, a discrete version of (A.12), is therefore defined as

$$ND_t = \sum_{i=1}^{N_t} w_{it} \sum_{s=1}^T y_{is}^{jit} x_s / T_i^{jit}. \quad (\text{A.13})$$

B Aggregate Quantification Techniques: A Review

Consider a survey that asks a sample of firms, for example, whether output growth x_t was “down”, “same” or “up” relative to the previous period. Since the proportion of respondents who replied “down”, “same” or “up” sum to unity the survey contains two pieces of independent information at time t .²⁴ Let U_t and D_t denote the proportion of firms that reported an output rise and fall.

Although quantification of categorical survey responses is to some extent arbitrary, since survey responses are a firm’s subjective assessment of the expected or actual behaviour of x_t , at the aggregate level quantitative measures of the expected or observed movement of x_t can be derived given certain assumptions. In this appendix three alternative methods of quantification are reviewed:

- the probability approach of Carlson and Parkin (1975);
- the regression approach of Pesaran (1984, 1987);
- the reverse-regression approach of Cunningham, Smith and Weale (1998) and Mitchell, Smith and Weale (2002a).

Although motivated in different ways, the three approaches are shown to share a common foundation. Our discussion compares the latter two methods to the probability approach.²⁵

B.1 The Probability Approach

This approach was first used by Theil (1952) to motivate the use of the “balance statistic” $U_t - D_t$ [see Anderson (1952)] as a method of quantification. The balance statistic, up to a scalar factor, provides an accurate measure of *average* output growth x_t if the percentage change in output of firms reporting a fall and the percentage change for firms reporting a rise are constant over time. The probability approach relaxes this restrictive assumption.

The probability method of quantification assumes that the response of firm i concerning economy-wide manufacturing output growth x_t is derived from a subjective probability density function for x_t , $f_i(\cdot|i)$, which may differ in form across firms and is conditional on information available to firm i at time t ; the dependence of $f_i(\cdot|i)$ on t is suppressed in the discussion.

The responses of firm i are classified as follows. Let $x_{it} = \int x f_i(x|i) dx$ denote the mean of $f_i(\cdot|i)$.

- “up” is observed if $x_{it} \geq b_{it}$;

²⁴The number of firms answering “not applicable” tends to be very small and is ignored here.

²⁵The exposition draws on Pesaran (1987) and Mitchell, Smith and Weale (2002a). For alternative reviews and extensions of the probability and regression approaches, see Wren-Lewis (1985) and Smith and McAleer (1995).

- “down” is observed if $x_{it} \leq -a_{it}$;
- “same” is observed if $-a_{it} < x_{it} < b_{it}$,

where the threshold parameters a_{it} and b_{it} are both positive.

Assume that firms are independent and that the structure of $f_i(\cdot|i)$ is the same and known for all firms; that is, $f_i(\cdot|i) = f(\cdot|i)$. Consequently, $x_{it} = \int x f(x|i) dx$ can be regarded as an independent draw from an aggregate density $f(x) = \int f(x|i) F(di)$, where $F(\cdot)$ denotes the distribution function of firms i ; the density $f(\cdot)$ is conditional on aggregate information available to all firms at time t , the dependence on which is again suppressed. Assume $f(\cdot)$ has mean x_t .

Furthermore, if the response thresholds are symmetric and are fixed both across firms i and time t , that is, $a_{it} = b_{it} = \lambda$, then

$$D_t \xrightarrow{p} P(x_{it} \leq -\lambda) = F_t(-\lambda), \quad (\text{B.1})$$

$$U_t \xrightarrow{p} P(x_{it} \geq \lambda) = 1 - F_t(\lambda), \quad (\text{B.2})$$

where $F_t(\cdot)$ is the cumulative distribution function obtained from $f(\cdot)$ where, now, we indicate explicitly the dependence on time t . Then, as x_{it} is an unbiased predictor for x_t , we can estimate x_t given a particular value for λ and a specific form for the aggregate distribution function $F_t(\cdot)$.

B.1.1 Carlson and Parkin’s Method

The traditional approach of Carlson and Parkin (1975) assumes that $f(\cdot)$ is a normal density function with mean x_t and variance σ_t ; alternative densities $f(\cdot)$ may be also considered; see Batchelor (1981) and Mitchell (2002).

From (B.1) and (B.2), the estimator for x_t is given as the solution to the equations

$$D_t = \Phi\left(\frac{-\lambda - \hat{x}_t}{\hat{\sigma}_t}\right), \quad (\text{B.3})$$

$$1 - U_t = \Phi\left(\frac{\lambda - \hat{x}_t}{\hat{\sigma}_t}\right), \quad (\text{B.4})$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Using (B.3) and (B.4) to solve for \hat{x}_t and $\hat{\sigma}_t$,

$$\hat{\sigma}_t = \frac{2\lambda}{\Phi^{-1}(1 - U_t) - \Phi^{-1}(D_t)}, \quad (\text{B.5})$$

where $\Phi^{-1}(\cdot)$ denotes the inverse standard normal cumulative distribution function. Thus,

$$\hat{x}_t = \lambda \left(\frac{\Phi^{-1}(1 - U_t) + \Phi^{-1}(D_t)}{\Phi^{-1}(1 - U_t) - \Phi^{-1}(D_t)} \right), \quad (\text{B.6})$$

which leaves only λ undetermined. In the literature λ has been calculated in various ways. Carlson and Parkin assume unbiasedness over the sample period, $t = 1, \dots, T$; that is, λ is estimated as

$$\hat{\lambda} = \left(\sum_{t=1}^T x_t \right) / \sum_{t=1}^T \left(\frac{\Phi^{-1}(1 - U_t) + \Phi^{-1}(D_t)}{\Phi^{-1}(1 - U_t) - \Phi^{-1}(D_t)} \right). \quad (\text{B.7})$$

For alternative approaches, see *inter alia* Batchelor (1981,1982), Pesaran (1984), and Wren-Lewis (1985). Since λ is constant over time, its rôle is merely to scale \hat{x}_t .

B.2 The Regression Approach

Let aggregate output x_t be a weighted average of firms' output x_{it} , ($i = 1, \dots, N_t$),

$$x_t = \sum_{i=1}^{N_t} w_i x_{it}, \quad (\text{B.8})$$

where w_i is the weight assigned to firm i . Assuming (B.8) holds for the sample of firms under consideration, and categorising firms according to whether they reported an “up” or a “down”, (B.8) can be rewritten as

$$x_t = \sum_{i=1}^{N_t} w_i^+ x_{it}^+ + \sum_{i=1}^{N_t} w_i^- x_{it}^- \quad (\text{B.9})$$

where x_{it}^+ is x_{it} if firm i reports an “up” and 0 otherwise, likewise, x_{it}^- equals x_{it} if firm i reports a “down” and 0 otherwise and w_i^+ and w_i^- the associated weights. The survey does not provide exact quantitative information on x_{it}^+ and x_{it}^- . Following Anderson, if, up to a mean zero disturbance ξ_{it} , $x_{it}^+ = \alpha$ and $x_{it}^- = -\beta$, $\alpha, \beta > 0$, then

$$x_t = \alpha \sum_{i=1}^{N_t} w_i^+ - \beta \sum_{i=1}^{N_t} w_i^- + \xi_t \quad (\text{B.10})$$

$$= \alpha U_t - \beta D_t + \xi_t, \quad (\text{B.11})$$

where $\xi_t = \sum_{i=1}^{N_t} w_i \xi_{it}$ and U_t and D_t are the (appropriately weighted) proportions of firms that reported an output rise and fall respectively. The unknown parameters α and β can be estimated *via* a linear (or non-linear) regression of x_t on U_t and D_t .²⁶ The fitted values from this estimated regression then provide the quantified retrospective survey response estimator for x_t . To ensure the fitted values are unbiased estimates for x_t , an intercept is also included in the regression to allow for the possibility that ξ_t has a time-invariant non-zero mean.

B.2.1 Relating the Regression Approach to the Probability Approach

Suppose that x_{it} is a random draw from a uniform density function $f(\cdot)$ with mean x_t and range $2q$, $q > 0$; that is,

$$\begin{aligned} f(x) &= (2q)^{-1} \text{ if } x_t - q \leq x \leq x_t + q, \\ &= 0 \text{ otherwise,} \end{aligned} \quad (\text{B.12})$$

²⁶For periods of rising and variable changes in x_t Pesaran extends this basic Anderson model to allow for an asymmetric relationship between x_t and x_{it} .

with corresponding cumulative distribution function

$$\begin{aligned} F_t(x) &= (2q)^{-1}[x - (x_t - q)] \text{ if } x_t - q \leq x \leq x_t + q \\ &= 0 \text{ if } x < x_t - q \\ &= 1 \text{ if } x > x_t + q. \end{aligned} \quad (\text{B.13})$$

From (B.2) and (B.1),

$$U_t = \frac{q + \hat{x}_t - \lambda}{2q}, \quad (\text{B.14})$$

$$D_t = \frac{q - \hat{x}_t - \lambda}{2q}, \quad (\text{B.15})$$

An estimate of output growth x_t may then be written as a function of the balance statistic; *viz.*

$$\hat{x}_t = q(U_t - D_t), \quad (\text{B.16})$$

which provides an alternative justification for the use of the balance statistic.

A generalisation of (B.16) is obtained by relaxing the assumption that the “no change” interval is symmetric; that is, replace $(-\lambda, \lambda)$ by $(-a, b)$. Hence, (B.14) and (B.15) become

$$U_t = \frac{q + \hat{x}_t - b}{2q}, \quad (\text{B.17})$$

$$D_t = \frac{q - \hat{x}_t - a}{2q}. \quad (\text{B.18})$$

Then the estimator for x_t is

$$\hat{x}_t = \alpha U_t - \beta D_t, \quad (\text{B.19})$$

which is equivalent to the estimator for x_t in (B.11) based on U_t and D_t for the single time period t , where the two scaling parameters are defined as

$$\alpha = \frac{2q(q - a)}{2q - a - b}, \quad \beta = \frac{2q(q - b)}{2q - a - b}. \quad (\text{B.20})$$

B.3 The Reverse-Regression Approach

Cunningham, Smith and Weale (1998) and Mitchell, Smith and Weale (2002a) relate survey responses to official data by relating the proportions of firms reporting rises and falls to the official data.²⁷ Under the assumption that (after revisions) official data offer unbiased estimates of the state of the economy this avoids biases caused by measurement error in the data.

²⁷We follow the approach of Mitchell, Smith and Weale (2002a) that modifies the original method of Cunningham, Smith and Weale (1998).

Following (A.1) let the categorical survey response of firm i at time t be determined by the firm-specific unobserved continuous random variable y_{it} which is related to economy-wide manufacturing output growth x_t through the linear representation

$$y_{it} = x_t + \eta_{it} + \varepsilon_{it}. \quad (\text{B.21})$$

The retrospective survey data provide firm level categorical information on the individual-specific random variable y_{it} via the discrete random variable y_{it}^j , $j = 1, 2, 3$, where

$$y_{it}^j = 1 \text{ if } c_{j-1} < y_{it} \leq c_j \text{ and } 0 \text{ otherwise,} \quad (\text{B.22})$$

where $c_0 = -\infty$ and $c_3 = \infty$, $j = 1, 2, 3$ with the intervals (c_0, c_1) , (c_1, c_2) and (c_2, c_3) corresponding to “down”, “same” and “up” respectively. Note that the thresholds c_j are invariant with respect to firm i and time t . Defined in terms of the error terms in (B.21), the observation rule (B.22) becomes

$$y_{it}^j = 1 \text{ if } c_{j-1} - x_t < \eta_{it} + \varepsilon_{it} \leq c_j - x_t \text{ and } 0 \text{ otherwise.} \quad (\text{B.23})$$

A probabilistic foundation may be given to the observation rule (B.23) by letting the scaled error terms $\{\sigma(\eta_{it} + \varepsilon_{it})\}$, $\sigma > 0$, possess a common and known cumulative distribution function $F(\cdot)$, $i = 1, \dots, N_t$, which is parameter free and assumed time-invariant. Then,

$$P(y_{it}^j = 1|x_t) = F(\mu_j - \sigma x_t) - F(\mu_{j-1} - \sigma x_t), \quad (\text{B.24})$$

where $\mu_j = \sigma c_j$, $j = 1, 2, 3$.

B.3.1 Motivating the Regression Formulation

Let the survey proportion of firms that give response j at time t be denoted by $P_t^j = \sum_{i=1}^{N_t} y_{it}^j / N_t$, $j = 1, 2, 3$. As $P_{jt} = P(y_{it}^j = 1|x_t) = F(\mu_j - \sigma x_t) - F(\mu_{j-1} - \sigma x_t)$, $E(P_t^j|x_t) = P_{jt}$. If we further assume that $F(\cdot)$ is symmetric, then $P_{1t} = F(\mu_1 - \sigma x_t)$ and $P_{3t} = F(-(\mu_2 - \sigma x_t))$. Hence, we may define the non-linear regressions

$$\begin{aligned} P_t^1 &= D_t = F(\mu_1 - \sigma x_t) + \xi_t^1, \\ P_t^3 &= U_t = F(-(\mu_2 - \sigma x_t)) + \xi_t^3. \end{aligned} \quad (\text{B.25})$$

Assuming that the survey responses of firms are independent given x_t ,

$$N_t^{1/2} \begin{pmatrix} \xi_t^1 \\ \xi_t^3 \end{pmatrix} \xrightarrow{d} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} F_t^1(1 - F_t^1) & -F_t^1 F_t^3 \\ -F_t^1 F_t^3 & F_t^3(1 - F_t^3) \end{pmatrix} \right), \quad (\text{B.26})$$

where $F_t^1 = F(\mu_1 - \sigma x_t)$ and $F_t^3 = F(-(\mu_2 - \sigma x_t))$. Restricting attention to categories $j = 1$ and $j = 3$ only results in no loss of information since $\sum_{j=1}^3 P_t^j = 1$.

If $F(\cdot)$ is strictly monotonic, the non-linear regressions (B.25) may be simplified by taking Taylor series approximations to $F^{-1}(D_t)$ and $F^{-1}(U_t)$ about $F(\mu_1 - \sigma x_t)$ and $F(-(\mu_2 - \sigma x_t))$ respectively yielding the *asymptotic* ($N_t \rightarrow \infty$) linear regression models

$$\begin{aligned} F^{-1}(D_t) &= \mu_1 - \sigma x_t + u_t^1, \\ F^{-1}(U_t) &= -\mu_2 + \sigma x_t + u_t^3, \end{aligned} \tag{B.27}$$

where

$$\begin{aligned} u_t^1 &= (f_t^1)^{-1} \xi_t^1 + o_p(N_t^{-1}), \\ u_t^3 &= (f_t^3)^{-1} \xi_{t,3} + o_p(N_t^{-1}), \end{aligned} \tag{B.28}$$

and $f_t^1 = f(\mu_1 - \sigma x_t)$, $f_t^3 = f(-(\mu_2 - \sigma x_t))$ and the density function $f(z) = dF(z)/dz$.

Since x_t is observed, feasible and asymptotically efficient estimation of (B.27) is achieved by generalised least squares (or minimum chi-squared) estimation given the structure of the variance-covariance matrix of u_t^1 and u_t^3 .

B.3.2 Estimation of x_t

Estimates of the official (economy-wide) macroeconomic data x_t may be derived from the estimated regressions. Consider the inverse regression model (B.27) and let

$$\hat{x}_t^1 = \frac{\hat{\mu}_1 - F^{-1}(D_t)}{\hat{\sigma}}, \quad \hat{x}_t^3 = \frac{\hat{\mu}_2 + F^{-1}(U_t)}{\hat{\sigma}}. \tag{B.29}$$

where $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\sigma}$ denote the coefficient estimates. Both \hat{x}_t^1 and \hat{x}_t^3 are consistent estimators of x_t . A reconciled estimator for x_t is obtained using the variance-covariance matrix of \hat{x}_t^1 and \hat{x}_t^3 [see Cunningham, Smith and Weale (1998) and Stone, Champernowne and Meade (1942)]. Note that when there is a poor statistical relationship between the survey proportions and x_t , σ tends to zero and the implied indicator becomes very volatile; see (B.29).

B.3.3 Relating the Reverse-Regression Approach to the Probability Approach

Let $F_t(x) = F((x - x_t)/\sigma_t)$ with $F(\cdot)$ symmetric. From (B.1) and (B.2) with an asymmetric interval for “same” $(-a, b)$, cf. (B.3) and (B.4), equate

$$1 - U_t = F\left(\frac{b - \hat{x}_t}{\hat{\sigma}_t}\right), \tag{B.30}$$

$$D_t = F\left(\frac{-a - \hat{x}_t}{\hat{\sigma}_t}\right). \tag{B.31}$$

From the symmetry of $F(\cdot)$,

$$U_t = F\left(\frac{-b + \hat{x}_t}{\hat{\sigma}_t}\right). \tag{B.32}$$

Hence,

$$F^{-1}(U_t) = \frac{-b + \hat{x}_t}{\hat{\sigma}_t}, \tag{B.33}$$

$$F^{-1}(D_t) = \frac{-a - \hat{x}_t}{\hat{\sigma}_t}. \tag{B.34}$$

Therefore, in comparison with (B.27), $\mu_1 = -a/\sigma_t$, $\mu_2 = b/\sigma_t$ and $\sigma = 1/\sigma_t$.

C The disaggregate indicator using prospective survey data

This appendix details how the disaggregate indicator can be operationalised when prospective survey data are considered. This involves an extension of the theory behind the Mitchell, Smith and Weale (2002a) estimator. Details are provided below; for the convenience of the reader they are taken directly from Mitchell, Smith and Weale (2004).

C.1 Firm-Level Quantification of Prospective Survey Data

We consider a survey that asks a sample of N_t manufacturing firms near the end of period $(t - 1)$ whether their output growth is expected to rise, not change or fall over period t compared to period $t - 1$. Our method of quantification of the prospective survey responses follows Mitchell, Smith and Weale (2002a) in postulating an underlying relationship between firm specific output growth and the official data for aggregate output growth and then extends their approach to the case of forward-looking or prospective survey responses. Our approach is influenced by the fact that the number of respondents varies from period to period, although a reasonable number respond in the majority of the periods.

The categorical responses in the survey are assumed to relate to observed official data for economy-wide manufacturing output growth x_t in the following manner. Let the actual output growth of firm i at time t , y_{it} , ($i = 1, \dots, N_t$), which may be known to firm i but is assumed unknown to the econometrician, depend on x_t according to the conditional linear model

$$y_{it} = x_t + \eta_{it} + \varepsilon_{it}, \quad (\text{C.1})$$

($t = 1, \dots, T$), where η_{it} is the difference between y_{it} and x_t anticipated by firm i , reflecting information private to firm i at time t that is not observed by the econometrician. This information may reflect firm or industry level influences. The random variable ε_{it} captures the component of firm-specific output growth y_{it} unanticipated by both firm i and the econometrician at time t . That is, $E(y_{it}|\Omega_t^i) = x_{it} = x_t + \eta_{it}$, where Ω_t^i comprises information available to firm i at time t and includes x_t . In the following analysis it is further assumed that output growth x_t is a stationary variable, an assumption supported by tests for a unit root in the level series of manufacturing output.

At the end of period $(t-1)$ firm i makes a prediction, y_{it}^* , of y_{it} based on macroeconomic information available to the firm, the relevant individual specific information set Ω_{t-1}^i , $i = 1, \dots, N_t$. If this prediction is formed rationally

$$y_{it}^* = E\{y_{it}|\Omega_{t-1}^i\} = x_t^* + \eta_{it}, \quad (\text{C.2})$$

where $x_t^* = E\{x_t|\Omega_{t-1}\}$ is the economy wide rational expectation of x_t , Ω_{t-1} is the macroeconomic information set available to all firms, and

$$x_t = x_t^* + \zeta_t, \quad (\text{C.3})$$

where ζ_t is a white-noise macroeconomic shock unanticipated by firms such that $E\{\zeta_t|\Omega_{t-1}\} = E\{\zeta_t|\Omega_{t-1}^i\} = 0$.

C.2 The relationship between the prospective survey data and the official data

We may re-express (C.2) as

$$y_{it}^* = \alpha_i + \beta_i x_t^* + \varepsilon_{it}^*, \quad (\text{C.4})$$

where α_i and β_i are firm-specific time-invariant coefficients that can be expressed in terms of (C.4) by defining $\eta_{it} = \alpha_i + (\beta_i - 1)x_t^* + \varepsilon_{it}^*$, ($i = 1, \dots, N_t$, $t = 1, \dots, T$), where ε_{it}^* is a mean zero random variable.²⁸ Substitution from (C.3) delivers

$$y_{it}^* = \alpha_i + \beta_i(x_t + \zeta_t) + \varepsilon_{it}^*, \quad (\text{C.5})$$

$$y_{it}^* = \alpha_i + \beta_i x_t + \psi_{it}, \quad (\text{C.6})$$

where $\psi_{it} = \beta_i \zeta_t + \varepsilon_{it}^*$. Estimation of (C.6) needs to take account of the endogeneity of x_t . We follow Smith and Blundell (1986) and Newey (1987) by supplementing (C.6) with the process assumed to govern the determination of x_t . This yields a two-equation simultaneous model which motivates an instrumental variables type estimator that corrects for the endogeneity.

We assume that an auto-regressive process governs the determination of x_t . Without loss of generality let us assume that x_t , measured relative to its mean, follows a first-order auto-regressive process²⁹

$$x_t = \lambda x_{t-1} + u_t, \quad (\text{C.7})$$

($t = 1, \dots, T$), where $|\lambda| < 1$ to ensure stationarity of output growth and u_t is an *i.i.d.* mean zero disturbance.

ψ_{it} and u_t follow a bivariate distribution; assume that conditionally on u_t we can write their dependence in the form

$$\psi_{it} = \rho_i u_t + \nu_{it}, \quad (\text{C.8})$$

where ρ_i ($i = 1, \dots, N_t$) is a firm-specific parameter and ν_{it} is an *i.i.d.* mean zero logistic disturbance distributed independently of u_t . Substitution in (C.6) generates the conditional model

$$y_{it}^* = \alpha_i + \beta_i x_t + \rho_i u_t + \nu_{it}. \quad (\text{C.9})$$

²⁸Furthermore, $E(y_{it}|\Omega_{t-1}) = \alpha + \beta x_t^*$ where $E(\alpha_i|\Omega_{t-1}) = \alpha$, $E(\beta_i|\Omega_{t-1}) = \beta$ and Ω_{t-1} comprises information available to *all* firms at time ($t - 1$). Let z_{it} denote (the level of) expected output of firm i at time t , so that $y_{it}^* = \frac{z_{it} - z_{it-1}}{z_{it-1}}$. From (C.4), $\sum_{i=1}^{N_t} \Delta z_{it} = \sum_{i=1}^{N_t} z_{it-1} \alpha_i + \sum_{i=1}^{N_t} z_{it-1} \beta_i x_t^* + \sum_{i=1}^{N_t} z_{it-1} \varepsilon_{it}^*$, after cross-multiplication and summation over $i = 1, \dots, N_t$, where Δ is the first difference operator. For coherence we require $\sum_{i=1}^{N_t} \Delta z_{it} / \sum_{i=1}^{N_t} z_{it-1} \xrightarrow{p} x_t^*$, $\sum_{i=1}^{N_t} z_{it-1} \alpha_i / \sum_{i=1}^{N_t} z_{it-1} \xrightarrow{p} 0$, $\sum_{i=1}^{N_t} z_{it-1} \beta_i / \sum_{i=1}^{N_t} z_{it-1} \xrightarrow{p} 1$ and $\sum_{i=1}^{N_t} z_{it-1} \varepsilon_{it}^* / \sum_{i=1}^{N_t} z_{it-1} \xrightarrow{p} 0$ ($N_t \rightarrow \infty$).

²⁹Additional auto-regressive terms can be included if necessary to render u_t serially uncorrelated.

Along the lines of Smith and Blundell (1986) and Newey (1987) we consider a two-step estimator. First, estimate (C.7) by Ordinary Least Squares. This yields a consistent estimate, $\hat{\lambda}$, for λ (as $T \rightarrow \infty$). Secondly, estimate the parameters in (C.9), i.e. α_i , β_i , μ_{ji} and ρ_i , by ordered logit substituting \hat{u}_t for u_t , where $\hat{u}_t = x_t - \hat{\lambda}x_{t-1}$. This is achieved as follows.

Anticipated growth y_{it}^* of firm i at time t is unobserved but the survey at time $(t-1)$ contains data corresponding to whether output growth is expected to rise, not change or fall in period t . To account for the ordinal nature of the responses, we use ordered discrete choice models [see Amemiya (1985), Ch.9] based on the latent regression (C.9). Define the indicator variables

$$y_{it-1}^j = 1 \text{ if } \mu_{(j-1)i} < y_{it}^* \leq \mu_{ji} \text{ and } 0 \text{ otherwise, } (j = 1, 2, 3), \quad (\text{C.10})$$

corresponding to “down”, “same” and “up”, respectively, where $\mu_{0i} = -\infty$, μ_{1i} , μ_{2i} and $\mu_{3i} = \infty$ are firm-specific threshold parameters. We assume that the error terms ν_{it} , ($t = 1, \dots, T$), are logistic with common cumulative distribution function (c.d.f.) $F(z) = [1 + \exp(-z)]^{-1}$, $-\infty < z < \infty$, ($i = 1, \dots, N_t$). The probabilistic foundation for the observation rule (C.10) is given by the conditional probability $P_{jit-1} \equiv P_i(j|x_t, i, \hat{u}_t)$ of observing the categorical response $y_{it-1}^j = 1$ for choice j at time $(t-1)$ given the values of x_t and \hat{u}_t and firm i

$$P_{jit-1} = F(\mu_{ji} - \alpha_i - \beta_i x_t - \rho_i \hat{u}_t) - F(\mu_{(j-1)i} - \alpha_i - \beta_i x_t - \rho_i \hat{u}_t), (j = 1, 2, 3). \quad (\text{C.11})$$

Assuming the errors ν_{it} are independently and identically distributed over time, the likelihood function for firm i is

$$L_i = \prod_{t=2}^T P_{1it-1}^{y_{it-1}^1} P_{2it-1}^{y_{it-1}^2} P_{3it-1}^{y_{it-1}^3}. \quad (\text{C.12})$$

Under the above assumptions, maximisation of (C.12) yields consistent estimates ($T \rightarrow \infty$) of α_i , β_i , ρ_i and μ_{ji} denoted by $\hat{\alpha}_i$, $\hat{\beta}_i$, $\hat{\rho}_i$ and $\hat{\mu}_{ji}$ respectively.

C.3 Inferring the Official Data in the Prospective Case

Given an ordered logit model for each firm i , an estimator for x_t may be inferred from the prospective survey data published at time $(t-1)$; following Mitchell, Smith and Weale (2002a) inference is based round Bayes' Theorem. In so doing the qualitative prospective survey data at time $(t-1)$ are converted into a quantitative indicator of expected growth in period t available to users of the survey at $(t-1)$. This indicator can then be evaluated as a one-step ahead forecast of x_t .

Let j_{it-1} , ($j_{it-1} = 1, 2, 3$), denote the prospective survey response of firm i published at time $(t-1)$, where 1, 2 and 3 correspond to “down”, “same” and “up”, respectively. Our initial interest centres on the conditional density $f(u_t|j_{t-1}, i, x_{t-1})$ for observing u_t given the survey response j for firm i at time $(t-1)$ and the time $(t-1)$ value of the official data. Let $f(u_t)$ denote the time-invariant probability density function (p.d.f.)

of u_t , where $f(u_t) = f(u_t|x_{t-1})$ since the shocks u_t that hit the aggregate economy are independent of x_{t-1} . This pd.f. is assumed normal with mean zero and variance $E(u_t^2)$.³⁰ Therefore, the conditional probability of observing response j_{t-1} for firm i is $P(j_{t-1}|i, x_{t-1}) = \int_{-\infty}^{\infty} P(j_{t-1}|x_{t-1}, i, u_t) f(u_t) du_t$. Bayes' Theorem states that

$$f(u_t|j_{t-1}, i, x_{t-1}) = \frac{P(j_{t-1}|x_{t-1}, i, u_t) f(u_t)}{P(j_{t-1}|i, x_{t-1})}. \quad (\text{C.13})$$

For firm i , the Bayes estimator (under squared error loss) for u_t given j and x_{t-1} is the mean of the posterior density $f(u_t|j_{t-1}, i, x_{t-1})$:

$$E(u_t|j_{t-1}, i, x_{t-1}) = \int_{-\infty}^{\infty} u_t f(u_t|j_{t-1}, i, x_{t-1}) du_t, \quad (\text{C.14})$$

which at time t takes one of three values depending on the observed sample response j_{it-1} of firm i at time $(t-1)$ and the lagged value of output growth x_{t-1} . Given $f(u_t)$, all of the above integrals may be calculated by numerical evaluation.

Estimators $\widehat{P}(j_{t-1}|x_{t-1}, i, \widehat{u}_t)$ for $P(j_{t-1}|x_{t-1}, i, u_t)$ and, thus, $\widehat{P}(j_{t-1}|i, x_{t-1})$ for $P(j_{t-1}|i, x_{t-1})$ are given by substitution of the estimators $\widehat{\alpha}_i$, $\widehat{\beta}_i$, $\widehat{\rho}_i$ and $\widehat{\mu}_{ji}$, ($j = 0, \dots, 3$), in (C.11) given that

$$\widehat{P}_{jit-1} = F(\widehat{\mu}_{ji} - \widehat{\alpha}_i - \widehat{\beta}_i x_t - \widehat{\rho}_i \widehat{u}_t) - F(\widehat{\mu}_{(j-1)i} - \widehat{\alpha}_i - \widehat{\beta}_i x_t - \widehat{\rho}_i \widehat{u}_t) \quad (\text{C.15})$$

$$\begin{aligned} &= F(\widehat{\mu}_{ji} - \widehat{\alpha}_i - \widehat{\beta}_i \widehat{\lambda} x_{t-1} - (\widehat{\beta}_i + \widehat{\rho}_i) \widehat{u}_t) \\ &\quad - F(\widehat{\mu}_{(j-1)i} - \widehat{\alpha}_i - \widehat{\beta}_i \widehat{\lambda} x_{t-1} - (\widehat{\beta}_i + \widehat{\rho}_i) \widehat{u}_t). \end{aligned} \quad (\text{C.16})$$

Hence, a feasible Bayes estimator $\widehat{E}(\widehat{u}_t|j_{t-1}, i, x_{t-1})$ may be obtained from (C.14) by numerical evaluation.

To create a disaggregate indicator D_t of economic activity at time $t-1$, from the law of iterated expectations the conditional expectation of x_t given all firms' survey responses j_{it-1} , ($i = 1, \dots, N_t$),

$$E(x_t|\{j_{it-1}\}_{i=1}^{N_t}, x_{t-1}) = \sum_{i=1}^{N_t} H_{it} E(x_t|j_{it-1}, i, x_{t-1}), \quad (\text{C.17})$$

where H_{it} is the exogenous sample probability of observing firm i at time t and, from (C.7),

$$E(x_t|j_{it-1}, i, x_{t-1}) = \lambda x_{t-1} + E(u_t|j_{it-1}, i, x_{t-1}). \quad (\text{C.18})$$

Hence, assuming firms are independent, we define the parametric indicator

$$\widehat{x}_t^D = \sum_{i=1}^{N_t} w_{it} \widehat{E}(x_t|j_{it-1}, i, x_{t-1}), \quad (\text{C.19})$$

where $w_{it} > 0$ is the weight given to firm i at time t and $\sum_{i=1}^{N_t} w_{it} = 1$. If firms constitute a random sample, then equal weights are appropriate since all firms are equally likely in

³⁰These assumptions can be tested by empirical tests.

the sample. However, if firms are drawn according to some stratified sampling process, then the weights w_{it} should reflect stratum weights; for example, if strata are defined by firm size, then firms should be size-weighted.

An alternative non-parametric disaggregate indicator \hat{x}_t^{ND} for the conditional expectation $E(x_t|\{j_{it-1}\}_{i=1}^{N_t})$ may be based on the conditional empirical distribution function. Define the indicator function $I(x_t \leq x, j_{it-1} = j|i) = 1$ if $x_t \leq x$ and $j_{it-1} = j$ and 0 otherwise, ($j = 1, 2, 3$). Let $T_i^j = \sum_{s=2}^T y_{i,s-1}^j$ which is the number of times firm i gives response j in the survey; hence, T_i^j/T is the sample proportion of responses j for firm i , ($j = 1, 2, 3$). The conditional empirical distribution function of x_t given response j for firm i is given by $\hat{F}(x|j, i) = \sum_{t=1}^T I(x_t \leq x, j_{it-1} = j|i)/T_i^j$, ($j = 1, 2, 3$), which assigns equal weight to each sample value. As $T \rightarrow \infty$ and, thus, $T_i^j \rightarrow \infty$, $T_i^j/T \xrightarrow{p} P(j|i)$ and $\sum_{t=1}^T I(x_t \leq x, j_{it-1} = j|i)/T \xrightarrow{p} F(x, j|i)$ if, given firm i , x_t and j_{it-1} may be regarded as stationary random variables with joint conditional c.d.f. $F(x, j|i)$. Hence, $\hat{F}(x|j, i) \xrightarrow{p} F(x|j, i) = F(x, j|i)/P(j|i)$, the conditional c.d.f. of x_t given response j and firm i . Therefore, the mean of $\hat{F}(x|j, i)$, $\sum_{s=2}^T y_{i,s-1}^j x_s / T_i^j$, is a consistent estimator for $E(x_t|j, i)$. A nonparametric disaggregate (\hat{x}_t^{ND}) indicator, a discrete cousin of (C.19), is therefore defined as

$$\hat{x}_t^{ND} = \sum_{i=1}^{N_t} w_{it} \sum_{s=2}^T y_{i,s-1}^j x_s / T_i^j. \quad (\text{C.20})$$

We note now that although the disaggregate indicators in practice have a good correlation with the official data they show much less volatility. Less volatility is observed because the scale is incorrect. One explanation for this is based on those firms whose prospective responses are poorly correlated with actual output growth. In the extreme case where responses are uncorrelated with output, the inclusion of these reduces the standard deviation of the indicator but does not affect its correlation with output growth. Excess smoothness of the disaggregate indicators can then be explained by the presence of firms in the sample of survey responses whose responses contain little signal about output growth and are essentially ‘noise’. To reconcile this incompatibility in volatility between the outturn and the indicators for manufacturing output growth, note that the outturn is the signal recovered from the survey data plus a residual error component. Rescaling the indicators through linear regression on the outturn is one simple method of obtaining an indicator which tracks output growth as closely as possible. The effects of this regression will also be taken into account in our subsequent out-of-sample analysis.

Specifically we align the disaggregate indicators with the official data by regressing the outturn x_t on the indicator as follows

$$x_t = \alpha_0 + \alpha_1 \hat{x}_t^k; \text{ for } k = D, ND. \quad (\text{C.21})$$

In fact, for the parametric indicator \hat{x}_t^D we consider the following unrestricted form of (C.21) that should better pick up the dynamic nature of x_t

$$x_t = \alpha_0 + \alpha_1^* x_{t-1} + \alpha_2^* \hat{u}_t^D, \quad (\text{C.22})$$

where $\widehat{u}_t^D = \sum_{i=1}^{N_t} w_{it} \widehat{E}(\widehat{u}_t | j_{it-1}, i, x_{t-1})$. Note that when $\alpha_1^* = \widehat{\lambda} \alpha_2^*$ only the magnitude of \widehat{x}_t^D is affected by the re-scaling.

C.4 Producing out-of-sample forecasts from the qualitative survey data

The section above provides a means of linking the prospective survey responses to official data on output growth. The techniques discussed can be used to quantify these forward-looking survey data in-sample. However, to be made operational out-of-sample we need to accommodate the fact that the official data for output growth are published with a lag.

From (C.7) we know

$$x_t = \widehat{\lambda} x_{t-1} + \widehat{u}_t. \quad (\text{C.23})$$

This implies that the expectation of x_t conditional on j_{t-1}, i and x_{t-1} is

$$\widehat{E}(x_t | j_{it-1}, i, x_{t-1}) = \widehat{\lambda} x_{t-1} + \widehat{E}(\widehat{u}_t | j_{it-1}, i, x_{t-1}). \quad (\text{C.24})$$

But out-of-sample we do not know x_{t-1} , although we do know j_{it-1} since the survey data are published ahead of the official data.³¹ We can however make use of the value generated as in (C.19). We denote this as $\widehat{x}_{t-2,t-1}^D$ to represent the fact that it is calculated only using information up to time $t-2$ using out-of-sample estimates of the logistic equations (C.9) whereas \widehat{x}_{t-1}^D was computed from estimates of the logistic equations including period $t-1$. We denote the density function of x_{t-1} conditional on $\widehat{x}_{t-2,t-1}^D$ as $g(x_{t-1} | \widehat{x}_{t-2,t-1}^D)$. We cannot estimate this from the individual logistic equations because the density function of the linear combination of these is unknown. Instead we explore the time-series relationship between x_{t-1} and $\widehat{x}_{t-2,t-1}^D$. We can test the hypothesis that $E(x_{t-1} - \widehat{x}_{t-2,t-1}^D)$ is zero and $x_{t-1} - \widehat{x}_{t-2,t-1}^D$ is normally distributed; if accepted we therefore assume the variance to be $E(x_{t-1} - \widehat{x}_{t-2,t-1}^D)^2$.

The second term on the right-hand-side of (C.24), using the forecast of x_{t-1} , is then given as

$$\widehat{E}(\widehat{u}_t | j_{it-1}, i, \widehat{x}_{t-2,t-1}^D) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \widehat{u}_t f(\widehat{u}_t | j_{it-1}, i, x_{t-1}) g(x_{t-1} | \widehat{x}_{t-2,t-1}^D) d\widehat{u}_t dx_{t-1}. \quad (\text{C.25})$$

An alternative approach would be to consider firm-specific forecasts for x_{t-1} instead of using our aggregate indicator $\widehat{x}_{t-2,t-1}^D$, and examine $\widehat{E}(x_{t-1} | j_{it-2}, i, x_{t-2})$; see (C.18). But given our aim of producing an aggregate forecast for x_t again it seems natural to base this on an aggregate forecast for x_{t-1} , $\widehat{x}_{t-2,t-1}^D$.

Following (C.22), the out-of-sample forecasts are re-scaled based on in-sample estimates of α_0 , α_1^* and α_2^* , denoted $\widehat{\alpha}_0$, $\widehat{\alpha}_1^*$ and $\widehat{\alpha}_2^*$, so that the re-scaled out-of-sample forecast, \widehat{x}_t^D , is given as

$$\widehat{x}_t^D = \widehat{\alpha}_0 + \widehat{\alpha}_1^* \widehat{x}_{t-2,t-1}^D + \widehat{\alpha}_2^* \sum_{i=1}^{N_t} w_{it} \widehat{E}(\widehat{u}_t | j_{it-1}, i, \widehat{x}_{t-2,t-1}^D). \quad (\text{C.26})$$

³¹For example, imagine it is January. While the survey has been published indicating what firms expect to happen in the first quarter, not only are official data for manufacturing output growth not available for this quarter but they are also not yet available for the last quarter of the previous year.

D Detailed tables of results for the retrospective survey data: in-sample

Table 6: Germany. 96 observations per firm. Performance of aggregate indicators

	Mean	S.D.	Corr.
Outturn	0.861	5.869	1.000
Full:CP	0.861	1.153	0.412
Full:P	0.861	2.443	0.416
Full:CSW	0.861	14.259	0.412
Full:CP weighted	0.861	2.661	0.397
Full:P size weighted	0.861	2.379	0.405
Full:CSW size weighted	0.861	14.287	0.410
Incl:CP	0.861	1.259	0.404
Incl:P	0.861	2.385	0.406
Incl:CSW	0.861	14.519	0.404
Excl:CP	0.861	1.126	0.397
Excl:P	0.861	2.346	0.400
Excl:CSW	0.861	14.402	0.407
Full: BAL	0.861	2.414	0.411

Notes: Full denotes full sample; Incl denotes included sample; Excl denotes excluded sample; CP denotes Carlson-Parkin; P denotes Pesaran; CSW denotes Cunningham, Smith and Weale; BAL denotes the balance statistic (unweighted) and size weighted denotes that the aggregate proportions are weighted according to 'firm size' as measured by the survey provider. Similar results are obtained for BAL in the size weighted case and are therefore not presented

Table 7: Germany. 96 observations per firm. Sample selection: p-values for $H_0: \pi_i - \pi_j = 0$

CP	0.959
P	0.965
CSW	0.983

Notes: p-values for the test considered in equation (1)

Table 8: Germany. 96 observations per firm. Disaggregate indicator performance

	Mean	S.D.	Corr.	RMSE	S.D. post regression
Outturn	0.861	5.869	1.000	0.000	5.869
N	0.861	0.174	0.710	4.117	4.166
NP	0.870	0.264	0.780	3.656	4.579
NP:Mode	0.889	0.242	0.782	3.640	4.591
N: R2	0.819	0.479	0.728	4.006	4.273
N:poly	0.823	0.459	0.652	4.431	3.827
NP R2	0.840	0.581	0.754	3.839	4.425
NP poly	0.843	0.544	0.702	4.163	4.119
N: RMSE	0.817	0.167	0.702	4.165	4.117
N: cor	0.818	0.328	0.715	4.089	4.193
NP: RMSE	0.827	0.251	0.780	3.657	4.577
NP: cor	-	-	-	5.907	-
N size weighted	0.824	0.203	0.604	4.659	3.543
NP size weighted	0.867	0.327	0.658	4.400	3.862

Notes: N denotes the parametric disaggregate indicator, denoted D_t in Appendix A; NP denotes the nonparametric disaggregate indicator, denoted ND_t in Appendix A; Mode indicates that the disaggregate indicator has been calculated replacing any missing values for a given firm by that firm's modal response; R2 denotes that the disaggregate indicator has been derived by weighting firm's quantified series according to the pseudo R^2 of the ordered logit model; poly denotes that the disaggregate indicator has been derived by weighting firm's quantified series according to the polyserial correlation between the categorical responses in the survey and the official data; RMSE and cor denote that the disaggregate indicator has been derived by weighting firm's quantified series according to its RMSE or correlation against the official data; size weighted denotes that the firm-level series are weighted according to the measure of firm-size given by the survey provider

Table 9: Germany. 96 observations per firm. Specification tests for ordered discrete choice models

Sp: Trad	0.927
Sp: Bonf	1.000
LM: M	0.474
LM: MS	0.523
LM: M B	0.831
LM: MS B	0.869

Notes: The table reports the proportion of times, across the firms, that the specification tests were not rejected. Sp denotes the joint specification test for no dynamics and exogeneity; LM: M denotes the Lagrange Multiplier test following Murphy (1996); LM: MS denotes the Lagrange Multiplier test following Machin and Stewart (1990); B denotes that the Bonferroni correction is used rather than traditional critical values

Table 10: Germany. 96 observations per firm. Proportion of times official data are not significant in firm-level discrete choice models

Trad	0.811
Bonf	0.996

Notes: Trad denotes that traditional critical values have been used in testing whether the official data are significant at a 95 per cent level; Bonf denotes that the Bonferroni correction has been used

Table 11: Germany. 96 observations per firm. Encompassing tests:
HAC standard errors

	PES:a1	PES:a2	PES:c	CP:a1	CP:a2	CP:c	CSW:a1	CSW:a2	CSW:c
NP: beta	1.211	-0.607	0.340	1.184	-0.547	0.312	1.236	-0.649	0.355
NP: se	0.108	0.207	0.435	0.103	0.200	0.439	0.114	0.212	0.426
NP: t	11.216	-2.927	0.782	11.513	-2.729	0.711	10.827	-3.053	0.834
N: beta	1.293	-0.702	0.352	1.263	-0.647	0.331	1.313	-0.724	0.354
N: se	0.131	0.252	0.432	0.125	0.240	0.438	0.137	0.262	0.426
N: t	9.899	-2.788	0.815	10.135	-2.699	0.755	9.565	-2.767	0.831

Notes: Pes:a1 refers to the parametric (N) or nonparametric (NP) disaggregate indicator in the regression of output growth on the disaggregate indicator and the Pesaran aggregate indicator; Pes:a2 refers to the Pesaran aggregate indicator in the regression of output growth on the disaggregate indicator and the Pesaran aggregate indicator disaggregate indicator and Pes:c refers to the constant in the regression. Similarly for CP and CSW, i.e. the Carlson Parkin and Cunningham, Smith and Weale aggregate indicators. NP: beta refers to the estimated coefficient; NP: se refers to the estimated standard error and NP: t refers to the t-ratio. Rows 2-4 therefore compare the informational content of the nonparametric disaggregate indicator against that of the three aggregate indicators in turn, while rows 5-7 consider the parametric disaggregate indicator

Table 12: Germany. 96 observations per firm. Encompassing tests
(cont.): p-values

	PES: Ha	PES: Hb	CP: Ha	CP: Hb	CSW: Ha	CSW: Hb
NP	0.00951	0.01802	0.00654	0.00000	0.00000	0.00000
N	0.02052	0.02595	0.02151	0.00000	0.00000	0.00000

Notes: Pes:Ha, or CP: Ha, or CSW: Ha, give the p-value of failing to reject Ha (the null hypothesis that the disaggregate indicator encompasses the aggregate indicator under consideration) while Hb refers to the alternative hypothesis that the aggregate indicator encompasses the disaggregate indicator; N refers to the parametric (N) and NP to the nonparametric (NP) disaggregate indicator

Table 13: Portugal. 80 observations per firm. Performance of aggregate indicators

	Mean	S.D.	Corr.
Outturn	1.883	11.387	1.000
Full:CP	1.883	4.108	0.073
Full:P	1.883	1.140	0.100
Full:CSW	1.883	94.471	0.121
Full:CP size weighted	1.883	4.566	0.081
Full:P size weighted	1.883	1.516	0.133
Full:CSW size weighted	1.883	59.311	0.192
Incl:CP	1.883	10.837	0.091
Incl:P	1.883	1.363	0.120
Incl:CSW	1.883	81.302	0.140
Excl:CP	1.883	2.010	-0.035
Excl:P	1.883	0.412	0.036
Excl:CSW	1.883	361.404	0.032
Full: BAL	1.883	0.830	0.073

Table 14: Portugal. 80 observations per firm. Sample selection: p-values for $H_0: \pi_i - \pi_j = 0$

CP	0.760
P	0.652
CSW	0.556

Table 15: Portugal. 80 observations per firm. Disaggregate indicator performance

	Mean	S.D.	Corr.	RMSE	S.D. post regression
Outturn	1.883	11.387	1.000	0.000	11.387
N	1.884	0.163	0.715	7.924	8.143
NP	1.869	0.327	0.735	7.686	8.371
NP:Mode	1.883	0.254	0.760	7.375	8.649
N: R2	1.640	0.361	0.737	7.658	8.396
N: poly	1.672	1.343	0.142	11.223	1.615
NP R2	1.593	0.534	0.778	7.116	8.865
NP poly	1.440	1.400	0.207	11.091	2.360
N: RMSE	1.661	0.158	0.635	8.757	7.232
N: cor	1.647	0.264	0.720	7.870	8.197
NP: RMSE	1.641	0.277	0.750	7.504	8.536
NP: cor	1.632	0.405	0.779	7.112	8.868
N size weighted	1.707	0.261	0.512	9.737	5.833
NP size weighted	1.880	0.400	0.652	8.597	7.423

Table 16: Portugal. 80 observations per firm. Specification tests for ordered discrete choice models

	Sp: Trad	0.924
Sp: Bonf	1.000	
LM: M	0.064	
LM: MS	0.189	
LM: M B	0.387	
LM: MS B	0.602	

Table 17: Portugal. 80 observations per firm. Proportion of times official data are not significant in firm-level discrete choice models

Trad	0.957
Bonf	0.998

Table 18: Portugal. 80 observations per firm. Encompassing tests:
HAC standard errors

	PES:a1	PES:a2	PES:c	CP:a1	CP:a2	CP:c	CSW:a1	CSW:a2	CSW:c
NP: beta	1.017	-2.179	4.072	1.051	-4.448	8.279	1.085	-6.602	12.270
NP: se	0.124	2.036	4.110	0.120	1.882	3.648	0.115	1.957	3.624
NP: t	8.219	-1.070	0.991	8.758	-2.364	2.270	9.460	-3.374	3.386
N: beta	1.032	-3.116	5.807	1.066	-4.992	9.275	1.088	-6.542	12.152
N: se	0.129	2.028	4.128	0.124	1.880	3.696	0.121	2.061	3.877
N: t	7.996	-1.537	1.407	8.575	-2.656	2.509	9.018	-3.174	3.134

Table 19: Portugal. 80 observations per firm. Encompassing tests
(cont.): p-values

	PES: Ha	PES: Hb	CP: Ha	CP: Hb	CSW: Ha	CSW: Hb
NP	0.54010	0.05341	0.00258	0.00000	0.00000	0.00000
N	0.25415	0.02250	0.00529	0.00000	0.00000	0.00000

Table 20: Sweden. 72 observations per firm. Performance of aggregate
indicators

	Mean	S.D.	Corr.
Outturn	1.742	7.835	1.000
Full:CP	1.742	2.601	0.215
Full:P	1.742	1.766	0.225
Full:CSW	1.742	33.803	0.232
Incl:CP	1.742	2.684	0.220
Incl:P	1.742	1.709	0.218
Incl:CSW	1.742	37.895	0.207
Excl:CP	1.742	2.683	0.197
Excl:P	1.742	1.866	0.238
Excl:CSW	1.742	30.282	0.259
Full: BAL	1.742	1.684	0.215

Table 21: Sweden. 72 observations per firm. Sample selection: p-values for $H_0: \pi_i - \pi_j = 0$

CP	0.905
P	0.918
CSW	0.788

Table 22: Sweden. 72 observations per firm. Disaggregate indicator performance

	Mean	S.D.	Corr.	RMSE	S.D. post regression
Outturn	1.742	7.835	1.000	0.000	7.835
N	1.739	0.215	0.619	6.123	4.847
NP	1.799	0.318	0.739	5.247	5.793
NP:Mode	1.758	0.281	0.745	5.201	5.835
N: R2	1.609	0.598	0.641	5.980	5.024
N: poly	1.615	0.774	0.478	6.847	3.743
NP R2	1.681	0.686	0.689	5.646	5.401
NP poly	1.674	0.836	0.506	6.721	3.966
N: RMSE	1.599	0.240	0.559	6.460	4.383
N: cor	1.602	0.417	0.597	6.254	4.675
NP: RMSE	1.652	0.335	0.686	5.671	5.374
NP: cor	1.648	0.462	0.713	5.467	5.584

Table 23: Sweden. 72 observations per firm. Specification tests for ordered discrete choice models

Sp: Trad	0.975
Sp: Bonf	1.000
LM: M	0.572
LM: MS	0.579
LM: M B	0.929
LM: MS B	0.938

Table 24: Sweden. 72 observations per firm. Proportion of times official data are not significant in firm-level discrete choice models

Trad	0.909
Bonf	0.998

Table 25: Sweden. 72 observations per firm. Encompassing tests: HAC standard errors

	PES:a1	PES:a2	PES:c	CP:a1	CP:a2	CP:c	CSW:a1	CSW:a2	CSW:c
NP: beta	1.222	-1.191	1.688	1.219	-1.503	2.236	1.194	-0.962	1.339
NP: se	0.099	0.270	0.665	0.108	0.395	0.898	0.094	0.242	0.702
NP: t	12.304	-4.402	2.538	11.337	-3.804	2.490	12.737	-3.975	1.907
N: beta	1.342	-1.331	1.721	1.464	-2.138	2.915	1.250	-0.926	1.178
N: se	0.159	0.335	0.748	0.187	0.593	1.115	0.149	0.302	0.791
N: t	8.438	-3.977	2.300	7.820	-3.604	2.615	8.398	-3.064	1.489

Table 26: Sweden. 72 observations per firm. Encompassing tests (cont.): p-values

	PES: Ha	PES: Hb	CP: Ha	CP: Hb	CSW: Ha	CSW: Hb
NP	0.00003	0.00038	0.00037	0.00000	0.00000	0.00000
N	0.00021	0.00112	0.00915	0.00000	0.00000	0.00000

Table 27: UK. 20 observations per firm. Performance of aggregate indicators. Results also presented using data up to 1997q3 only to facilitate comparison with size-weighted results

	Mean	S.D.	Corr.
Outturn	1.086	4.045	1.000
Full:CP	1.086	4.811	0.567
Full:P	1.086	2.371	0.586
Full:CSW	1.086	6.704	0.603
Incl:CP	1.086	4.959	0.545
Incl:P	1.086	2.302	0.569
Incl:CSW	1.086	6.944	0.583
Excl:CP	1.086	4.957	0.572
Excl:P	1.086	2.366	0.585
Excl:CSW	1.086	6.749	0.599
Full: BAL	1.086	2.360	0.583
Full: CP (97q3)	1.023	175.748	0.696
Full: CP size weighted (97q3)	1.023	88.501	0.648
Full: P (97q3)	1.023	2.978	0.734
Full: P size weighted (97q3)	1.023	2.752	0.678
Full: CSW (97q3)	1.023	5.574	0.728
Full: CSW size weighted (97q3)	1.023	6.244	0.650

Table 28: UK. 20 observations per firm. Sample selection: p-values for $H_0: \pi_i - \pi_j = 0$

CP	0.896
P	0.937
CSW	0.931

Table 29: UK. 20 observations per firm. Disaggregate indicator performance. Results also presented using data up to 1997q3 only to facilitate comparison with size-weighted results

	Mean	S.D.	Corr.	RMSE	S.D. post regression
Outturn	1.086	4.045	1.000	0.000	4.045
N	1.077	0.365	0.812	2.336	3.284
NP	1.040	0.547	0.910	1.658	3.681
NP:Mode	1.107	0.340	0.860	2.044	3.477
N: R2	0.670	0.591	0.827	2.249	3.345
N: poly	0.703	0.543	0.743	2.675	3.007
NP: R2	0.617	0.734	0.881	1.895	3.562
NP: poly	0.650	0.669	0.835	2.202	3.377
N: RMSE	0.708	0.282	0.704	2.841	2.847
N: cor	0.693	0.442	0.776	2.524	3.138
NP: RMSE	0.678	0.390	0.851	2.101	3.442
NP: cor	0.643	0.524	0.876	1.926	3.545
N: (to 97q3)	1.031	0.427	0.851	-	-
N: size weighted (to 97q3)	1.054	0.481	0.810	-	-
NP: (to 97q3)	1.009	0.640	0.923	-	-
NP: size weighted (to 97q3)	1.008	0.685	0.905	-	-

Table 30: UK. 20 observations per firm. Specification tests for ordered discrete choice models

Sp: Trad	0.919
Sp: Bonf	1.000
LM: M	0.488
LM: MS	0.488
LM: M B	0.964
LM: MS B	0.975

Table 31: UK. 20 observations per firm. Proportion of times official data are not significant in firm-level discrete choice models

Trad	0.808
Bonf	0.999

Table 32: UK. 20 observations per firm. Encompassing tests: HAC standard errors

	PES:a1	PES:a2	PES:c	CP:a1	CP:a2	CP:c	CSW:a1	CSW:a2	CSW:c
NP: beta	1.560	-0.814	0.260	1.410	-0.653	0.248	1.448	-0.693	0.251
NP: se	0.101	0.133	0.176	0.119	0.162	0.234	0.092	0.129	0.219
NP: t	15.506	-6.107	1.473	11.824	-4.020	1.058	15.723	-5.363	1.147
N: beta	1.608	-0.787	0.183	1.699	-0.961	0.268	1.675	-0.910	0.241
N: se	0.201	0.302	0.312	0.238	0.323	0.316	0.195	0.300	0.282
N: t	7.982	-2.602	0.586	7.151	-2.977	0.847	8.595	-3.037	0.852

Table 33: UK. 20 observations per firm. Encompassing tests (cont.): p-values

	PES: Ha	PES: Hb	CP: Ha	CP: Hb	CSW: Ha	CSW: Hb
NP	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
N	0.00792	0.00687	0.00086	0.00000	0.00000	0.00000