RESERVE BANK
OF NEW ZEALAND

DP2007/02

Nowcasting and Predicting Data Revisions in Real Time Using Qualitative Panel Survey Data

Troy Matheson, James Mitchell and Brian Silverstone

January 2007

JEL classification: C35, C42, C53, C80

www.rbnz.govt.nz/research/discusspapers/

Discussion Paper Series

DP2007/02

Nowcasting and Predicting Data Revisions in Real Time Using Qualitative Survey Data*

Troy Matheson, James Mitchell and Brian Silverstone †

Abstract

The qualitative responses that firms give to business survey questions regarding changes in their own output provide a real-time signal of official output changes. The most commonly-used method to produce an aggregate quantitative indicator from business survey responses – the net balance, or diffusion index – has changed little in 40 years. It focuses on the proportion of survey respondents replying "up", "the same" or "down". This paper investigates whether an improved real-time signal of official output data changes can be derived from a recently advanced method on the aggregation of survey data from panel responses. It also considers the ability of survey data to anticipate revisions to official output data. We find, in a New Zealand application, that exploiting the panel dimension to qualitative survey data gives a better in-sample signal about official data than traditional methods. This is achieved by giving a higher weight to firms whose answers have a close link to official data than to those whose experiences correspond only weakly or not at all. Out-of-sample, it is less clear it matters how survey data are quantified with simpler and more parsimonious methods hard to improve. It is clear, nevertheless, that survey data, exploited in some form, help to explain revisions to official data.

---

* The views expressed in this paper are those of the author(s) and do not necessarily reflect those of the Reserve Bank of New Zealand or the New Zealand Institute of Economic Research (NZIER). All errors and omissions are entirely ours. Access to the Quarterly Survey of Business Opinion (QSBO) data used in this study was provided by the NZIER under conditions that maintained full confidentiality. The authors are grateful for access to this data set.

† Troy Matheson, Economics Department, RBNZ, P.O.Box 2498, Wellington, New Zealand, Tel: +64-4-471-3859; Fax +64-4-473-1209; email: troy.matheson@rbnz.govt.nz. James Mitchell, National Institute of Economic and Social Research, 2 Dean Trench Street, Smith Square, London, SW1P 3HE, U.K. Tel: +44-207-654-1926. Fax: +44-207-654-1900; email: j.mitchell@niesr.ac.uk. Brian Silverstone, University of Waikato and Research Associate NZIER. Economics Department, Private Bag 3105, Hamilton, New Zealand. Tel: +64-7-838-4076; Fax: +64-7-838-4331; email: silver@waikato.ac.nz.

1 Introduction

The collection and publication of official data is subject to processing delays. As a result, considerable resources are devoted to the construction of nowcasts within central banks and policy institutions. Since qualitative surveys about the state of an economy are published with a much shorter lag than official data, they are often useful in the construction of 'nowcasts'. The gain in timeliness from surveys, however, sometimes involves a loss of accuracy. The nature of this trade-off is thus critical when deciding how much weight to give to particular nowcasts.

The net balance statistic (the difference between the proportion of optimists and the proportion of pessimists) is the most commonly used method of converting qualitative survey responses into a quantitative measure useful for nowcasting. In the academic literature, however, a wider range of measures have long been discussed. See, for example, the surveys by Driver and Urga (2004), Mitchell *et al* (2006), Nardo (2003), Pesaran and Weale (2006), and Smith and McAleer (1995). These alternative measures include the probability and regression approaches of Carlson and Parkin (1975) and Pesaran (1984), respectively. A common feature of these traditional approaches is that they use of aggregate rather than individual responses; typically the *proportions* of optimists and pessimists are used.

Mitchell *et al* (2006) (henceforth MSW) have argued that there is no reason to believe that working with aggregate survey data is the best way of constructing nowcasts. It may be that quantification that allows for heterogeneity among firms could exploit survey responses more efficiently than traditional approaches. The outcomes could include more accurate inferences about underlying official variables such as output. A key feature of MSW's indicator is that it gives more weight to firms whose answers have a closer link to official data than to those whose experiences correspond only weakly or not at all. Their approach – a development of Mitchell *et al* (2005b) applied to prospective survey data – can be seen as a variant of the forecast combination problem considered by Granger and Ramanathan (1984), although the form of the problem is different.

In this paper – involving an application to New Zealand – we undertake two main tasks. First, we compare the real-time signalling performance of the MSW indicator against four traditional quantification techniques: the net balance statistic, the probability method of Carlson and Parkin (1975), the regression approach of Pesaran (1984) and Pesaran (1987), and the reverse regression approach of Cunningham *et al* (1998). Out-of-sample, we consider a time series (autoregressive) benchmark, the random walk model. We also compare nowcasts from the above models with Reserve Bank of New Zealand (RBNZ) nowcasts. Our application represents the most comprehensive examination to-date of the merits of exploiting firm-level survey data when nowcasting. While MSW applied their method to UK data, they were constrained, by data availability, to out-of-sample experiments over just eight quarters. In addition they confined attention to near-final, rather than real-time, official data.

We offer advice to practitioners by identifying when and why the MSW indicator is likely to perform well relative to its competitors. Secondly, we test the relative usefulness of the MSW indicator in predicting revisions to official output data. Our work is similar in spirit to a long-term project initiated by the Federal Reserve Bank of Philadelphia, and the Federal Reserve System more widely, on the importance of data revisions (Stark 2002).

Section 2 introduces the Quarterly Survey of Business Opinion (QSBO) published by the New Zealand Institute of Economic Research (NZIER) and the implementation of the MSW indicator. Section 3 considers both traditional survey-based nowcasts and selected nowcasts produced by the Reserve Bank of New Zealand. Section 4 describes the real-time data set for official GDP and manufacturing GDP used in the out-of-sample experiments. This data set provides a genuine real-time assessment of the performance of the MSW indicators and its competitors. Sections 5 and 6 detail the in-sample and out-of-sample results. These include an out-of-sample assessment of the ability of survey data and the MSW indicator to predict revisions to official data. Our conclusions are in Section 7.

2 The QSBO and implementation of the MSW indicator

Since 1961, the New Zealand Institute of Economic Research (NZIER) has conducted a Quarterly Survey of Business Opinion (QSBO) of executives in the manufacturing, building, merchant and service sectors. All questions in the survey, but one, relate to the microeconomic experiences and outlook of firms and almost all involve qualitative responses (of the up/same/down type). As our interest is in the production of nowcasts for output growth, we consider the QSBO question that asks firms about their experience (up/same/down/not applicable) during the past three months (excluding seasonal variations) regarding own output. This retrospective question provides the basis for deriving timely indicators for GDP and manufacturing GDP growth, called, collectively, x_t. (The number answering "not applicable" is very small and is ignored in what follows).

2.1 The Mitchell, Smith and Weale indicator: a review

The MSW methodology assumes that the categorical responses in the survey are related to economy-wide and manufacturing output growth x_t in the following manner. Let the actual output growth of firm i at time t, y_{it}, $(i = 1,...,N_t)$, depend on x_t according to the linear model:

$$y_{it} = \alpha_i + \beta_i x_t + \varepsilon_{it}, \tag{1}$$

$(t = 1,\ldots,T)$, where α_i and β_i are firm-specific time-invariant coefficients. The error term ε_{it} captures the component of firm-specific output growth y_{it} unanticipated by both firm i and the econometrician at time t. More precisely, we assume the conditional linear specification $E(y_{it}|\Omega_t^i) = \alpha_i + \beta_i x_t$ where Ω_t^i comprises information available to firm i at time t and includes x_t. Hence, $E(\varepsilon_{it}|\Omega_t^i) = 0$ and ε_{it} is uncorrelated with x_t, rendering x_t weakly exogenous by assumption. The validation of this and other assumptions, such as the absence of dynamics in x_t, is a necessary concomitant in any empirical application. Indeed, the model (1) can be straightforwardly augmented to accommodate the endogeneity and dynamic dependence in x_t. In the following analysis it is further assumed that output growth x_t is stationary.[1]

Actual growth y_{it} of firm i at time t is unobserved. The survey, however, 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 based on the latent regression (1). Define the indicator variables

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

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 ε_{it}, for $i = 1,\ldots,N$ and $t = 1,\ldots,T$, are logistic with common cumulative distribution function (cdf) $F(z) = [1+\exp(-z)]^{-1}$, $-\infty < z < \infty$. The logistic distribution is similar in shape to the normal but has slightly heavier tails. This distribution is particularly convenient since it offers a closed form distribution function. The probabilistic foundation for the observation rule (2) is given by

[1] It is necessary that model (1) for firm-level growth y_{it} is coherent with the economy-wide outturn x_t. Let z_{it} denote (the level of) output of firm i at time t. From (1), after cross-multiplication and summation over $i = 1,\ldots,N_t$, $\sum_{i=1}^{N_t} \Delta z_{it} = \sum_{i=1}^{N_t} z_{it-1}\alpha_i + \sum_{i=1}^{N_t} z_{it-1}\beta_i x_t + \sum_{i=1}^{N_t} z_{it-1}\varepsilon_{it}$, where Δ is the first difference operator. Therefore, for coherency, we require that $\sum_{i=1}^{N_t} \Delta z_{it}/\sum_{i=1}^{N_t} z_{it-1} \xrightarrow{p} x_t$, $\sum_{i=1}^{N_t} z_{it-1}\alpha_i/\sum_{i=1}^{N_t} z_{it-1} \xrightarrow{p} 0$, $\sum_{i=1}^{N_t} z_{it-1}\beta_i/\sum_{i=1}^{N_t} z_{it-1} \xrightarrow{p} 1$ and $\sum_{i=1}^{N_t} z_{it-1}\varepsilon_{it}/\sum_{i=1}^{N_t} z_{it-1} \xrightarrow{p} 0$ $(N_t \to \infty)$.

the conditional probability $P_{jit} = P_i(j|x_t, i)$ of observing the categorical response $y_{it}^j = 1$ for choice j at time t given the value of x_t and firm i

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

As discrete choice models are only identified up to scale, including the intercept α_i in (1) necessitates setting, for example, the first threshold parameter μ_{1i} to zero to achieve identification. Consequently the decision probabilities (3) are invariant to multiplying (1) by an arbitrary constant. Assuming the errors ε_{it} are independently and identically distributed over time, the likelihood function for firm i is

$$L_i = \prod\nolimits_{t=1}^{T} P_{1it}^{y_{it}^1} P_{2it}^{y_{it}^2} P_{3it}^{y_{it}^3}. \tag{4}$$

Under the above assumptions, maximisation of (4) yields consistent estimates of α_i, β_i and μ_{ji}, denoted by $\hat{\alpha}_i$, $\hat{\beta}_i$ and $\hat{\mu}_{ji}$, respectively. In addition, assuming firms are randomly sampled and that the model (1) is correctly specified, the errors ε_{it} are also independently distributed over firms. If we let $\varepsilon_{.t}$ denote the N_t-vector of ε_{it}'s and let $\Sigma_t = Var(\varepsilon_{.t})$, this implies that Σ_t is diagonal. There is, in this case, no loss of efficiency in estimating the discrete choice models firm-by-firm rather than as a system.

2.2 Inferring the Official Data: the MSW Indicator

Given an ordered logit model for each firm i, $i = 1, \ldots, N_t$, an estimator for x_t may be inferred from these survey data. As survey data are usually published ahead of the official data, this provides the basis for early quantitative estimates of x_t.

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. We need to determine the density function of x_t conditional on the N_t firms' observed survey responses at time t, $\{j_{it}\}_{i=1}^{N_t}$. We denote this density function $f(x_t|\{j_{it}\}_{i=1}^{N_t})$.

Let $f(x_t)$ denote the prior density function of x_t. This density function can be conditioned on lagged values of x_t, say x_{t-1}, when x_t follows a dynamic process. Also as MSW explain, for each firm one can consider ordered logit models augmented with x_{t-1} when x_{t-1} is statistically significant in the firm-level model (1). The exposition below, without loss of generality, considers both the unconditional density for x_t and the case when x_{t-1} is statistically insignificant in the firm-level model.

Diagonality of Σ_t implies that conditional on x_t firms' categorical responses are independent across firms. That is, the joint conditional probability of observing the

N_t firms' categorical responses, $\{j_{it}\}_{i=1}^{N_t}$, is given as the product of their marginal probabilities $P(j_{it}|x_t, i)$:

$$P(\{j_{it}\}_{i=1}^{N_t}|x_t) = \prod_{i=1}^{N_t} P(j_{it}|x_t, i). \tag{5}$$

Therefore, the joint conditional probability of observing response j across firms i, $(i = 1, \ldots, N_t)$, is given as

$$P(\{j_{it}\}_{i=1}^{N_t}) = \int_{-\infty}^{\infty} \prod_{i=1}^{N_t} P(j_{it}|x_t, i) f(x_t) dx_t. \tag{6}$$

Bayes' Theorem states that:

$$f(x_t|\{j_{it}\}_{i=1}^{N_t}) = \frac{\prod_{i=1}^{N_t} P(j_{it}|x_t, i) f(x_t)}{P(\{j_{it}\}_{i=1}^{N_t})}. \tag{7}$$

MSW then define the indicator D_t. D_t is given as the Bayes estimator (under squared error loss) for x_t given $\{j_{it}\}_{i=1}^{N_t}$ which is the mean of the posterior density $f(x_t|\{j_{it}\}_{i=1}^{N_t})$:

$$D_t = E(x_t|\{j_{it}\}_{i=1}^{N_t}) = \int_{-\infty}^{\infty} x_t f(x_t|\{j_{it}\}_{i=1}^{N_t}) dx_t. \tag{8}$$

Given $f(x_t)$, all of the integrals above may be calculated by numerical evaluation. Estimators $\hat{P}(j_{it}|x_t, i)$ for $P(j_{it}|x_t, i)$ and, thus, $\hat{P}(j_{it}|i)$ for $P(j_{it}|i)$ are given by substitution of the estimators $\hat{\alpha}_i$, $\hat{\beta}_i$ and $\hat{\mu}_{ji}$, $(j = 0, ..., 3)$, in (3). Hence, a feasible Bayes estimator $D_t = \hat{E}(x_t|\{j_{it}\}_{i=1}^{N_t})$ may be obtained from (8) by numerical evaluation.

The MSW indicator D_t considers all firms' responses, $i = 1, \ldots, N_t$, simultaneously. It is designed to give more weight to firms whose answers have a close link to the official data than to those whose experiences correspond only weakly or not at all. This can be seen as a variant of the forecast combination problem addressed by Bates and Granger (1969) and Granger and Ramanathan (1984). There are reasons why some firms might be more useful indicators than others, ranging from the nature of the business they operate to the care they employ in completing the survey. It is intuitive, then, that the study of individual firms' performances provides valuable information lost in aggregation.

To illustrate this property consider $\beta_1 = 0$. This implies that firm 1's categorical survey responses offer no information about the official data. For this firm

$P(j_{it}|x_t, i) = P(j_{it}|i)$. (7) is then given as

$$f(x_t|\{j_{it}\}_{i=1}^{N_t}) = \frac{\prod_{i=2}^{N_t} P(j_{it}|x_t, i) f(x_t)}{P(\{j_{it}\}_{i=2}^{N_t})}, \tag{9}$$

implying firm 1 receives no weight in the indicator D_t.

2.3 QSBO and implementation of the MSW indicator

Our QSBO sample contains the survey responses of 2,512 firms over the period 1983Q3 to 2006Q1 (91 quarters). Of this number, 922 firms are in the manufacturing sector. There are, on average, 542 firms in the sample at time t with 20 or more time-series observations. Many observations are missing as firms do not always respond to consecutive surveys. This outcome has implications for the implementation of the MSW indicator. In particular, it prevents the construction of a panel data set with sufficient time-series observations across all firms for the estimation of (1), without assuming some homogeneity in behaviour across firms. Quantification based on (1) requires sufficient time-series observations for a given firm for reliable parameter estimation.

In our application, we consider 20 observations to be satisfactory.[2] These observations need not be consecutive. Hence, firms that do not respond to at least 20 surveys are dropped from the sample used to derive the MSW indicators. Since these firms are dropped, there is a danger that the sample selection could induce bias in the MSW indicator.[3] Ultimately, the MSW indicators are determined by how well they perform both in-sample and out-of-sample relative to traditional quantification methods using aggregate survey data.

We also consider an approach not covered in MSW and not requiring any firms to be removed. This involves a pooled reformulation of (1) that imposes homogeneity restrictions across firms: simply re-express (1) as $y_{it} = \alpha + \beta x_t + \zeta_{it}$. In other words, rather than estimating firm-level models relating the survey and official data, we estimate a common specification for all firms. While this has the attraction of using all available data, the estimates are inconsistent in the presence of heterogeneity across firms. Nevertheless, for completeness, we present results using this pooled discrete choice model, denoted MSW (pooled).

[2] This choice is somewhat arbitrary and warrants further investigation via Monte-Carlo experiments. In practice, the MSW indicator appears to behave similarly across a wide range of cut-off values.

[3] Mitchell *et al* (2005a) and Mitchell *et al* (2005b) propose simple tests for sample selection.

3 Competitors to MSW

3.1 Traditional survey-based nowcasts

Published survey data usually report just the aggregate proportion of respondents who reply “up”, “the same” or “down”, and/or the balance of opinion (“ups”minus “downs”). The four traditional quantification methods against which MSW will be compared all use aggregate survey proportions, namely, the balance statistic [BAL], the probability method of Carlson and Parkin (1975) [CP], the regression approach of Pesaran (1984) and Pesaran (1987) [PES], and the reverse-regression approach of Cunningham *et al* (1998) [CSW]. Out-of-sample, we also consider a time-series (auto-regressive) benchmark, the random walk model.

3.2 RBNZ nowcasts for GDP and manufacturing GDP

As part of the RBNZ’s quarterly forecasting round, ‘first-pass’ central projections for GDP growth are presented to the *Monetary Policy Committee* (MPC) for deliberation. Because estimates for a given quarter are revised with the arrival of additional information – just like official data – these estimates are produced on a rolling basis. In an attempt to match the timing a nowcast can be produced from QSBO, two ‘first-pass’ estimates are considered: a ‘first-pass’ estimate for GDP growth in quarter T using information available up to the end of the second month of quarter $T+1$ (denoted RBNZ1) and a ‘first-pass’ estimate for GDP in quarter T formed at the end of the second month of quarter T (denoted RBNZ2). This latter estimate involves a larger component of forecasting (a smaller information set) than the former. A QSBO-based nowcast can be formed at some time between these two estimates.

The RBNZ produces a nowcast for manufacturing GDP growth in quarter $T+1$ using information on hours paid in the manufacturing sector. Data on hours paid for quarter T are available at the beginning of the second month of quarter $T+1$ (around two weeks before the first-pass projections for GDP growth are finalised). The nowcast is computed using a regression of manufacturing GDP growth up to quarter $T-1$ on a constant and hours paid growth up to quarter $T-1$. In real-time, nowcasts are then obtained by using the quarter T value of hours paid to produce a nowcast for manufacturing GDP growth for quarter T. A QSBO nowcast could be available a month ahead of this RBNZ nowcast. We examine the fitted values from this regression estimated over the full-sample period when we evaluate the RBNZ nowcast for manufacturing GDP growth in-sample.

4 Real-time official data

To evaluate the MSW indicator and to form a view about its likely real-time performance, we conducted out-of-sample simulations. These exploit the real-time data-sets (triangles) for GDP and manufacturing GDP (excluding the primary food manufacturing sector) compiled by the RBNZ. See Sleeman (2006) for details on the real-time data. In-sample, we focus on the latest (final) vintage, as this is believed to provide the most accurate estimates of the true data.

Real-time data for manufacturing GDP (excluding the primary food manufacturing sector) are strictly quasi-real-time estimates. While real-time data are available for the manufacturing sector, they are not available for the primary food manufacturing sector. The data are computed as the difference between the real-time manufacturing data and the current estimates of primary food manufacturing. These quasi-real-time data are available from the 2001Q3 vintage. Each vintage contains data back to 1987Q2.[4]

4.1 The reference series: quarterly growth at an annual rate

There is always a question about the appropriate transformation of the official data. We consider nowcasting the quarterly growth rate of GDP and manufacturing GDP (at an annual rate). In our opinion, a comparison of the survey with the immediate past (the last quarter) is more informative about the survey's usefulness than comparison against the more distant past (such as the same quarter in the previous year). If we were, for example, to consider the annual growth rate as the reference series, then, when quantifying the survey data at quarter T, official data for the reference series would have been published for at least a couple of quarters in the year. Expressed differently, at time T hard data on the reference (official) series are known for part of the period to which the reference series relates. There is the fear, then, of using the survey data to tell us, in large part, what we already know.

The observation 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 two to three quarters and the fact that annual growth rates

[4] We use historical data from vintages post-2001Q3 to balance the panel from 2000Q1 to 2001Q2 – that is, we use the 2001Q3 historical data to fill the gaps. This, as we see below, facilitates the production of nowcasts in the out-of-sample experiments from the beginning of 2000. For completeness, we also considered the performance of the MSW indicator using annual growth rates. The results were similar to those presented below for the quarterly growth rate.

are smoother than quarterly growth rates, explains the higher correlation found between the retrospective survey data and annual output growth compared with quarterly growth.[5]

5 In-sample evaluation

Firm-level ordered logit models are estimated for the 950 firms (318 firms in the manufacturing sector) that replied to at least 20 surveys. Pooled models are also estimated on the entire panel of firms. There is clearly considerable heterogeneity across firms since a Wald test (fixed N_t) rejects the null hypothesis $\beta_i = \beta$ for all i with a p-value of 0.00.

As described above, to compute the MSW indicator all that remains is to specify the prior density for the official data, $f(.)$. We assume a normal density with mean and variance set equal to the sample moments for the official data x_t, where we assume that x_t is stationary. When $f(.)$ is specified out-of-sample, we continue to assume normality but estimate its moments using only the information that was available at the time.

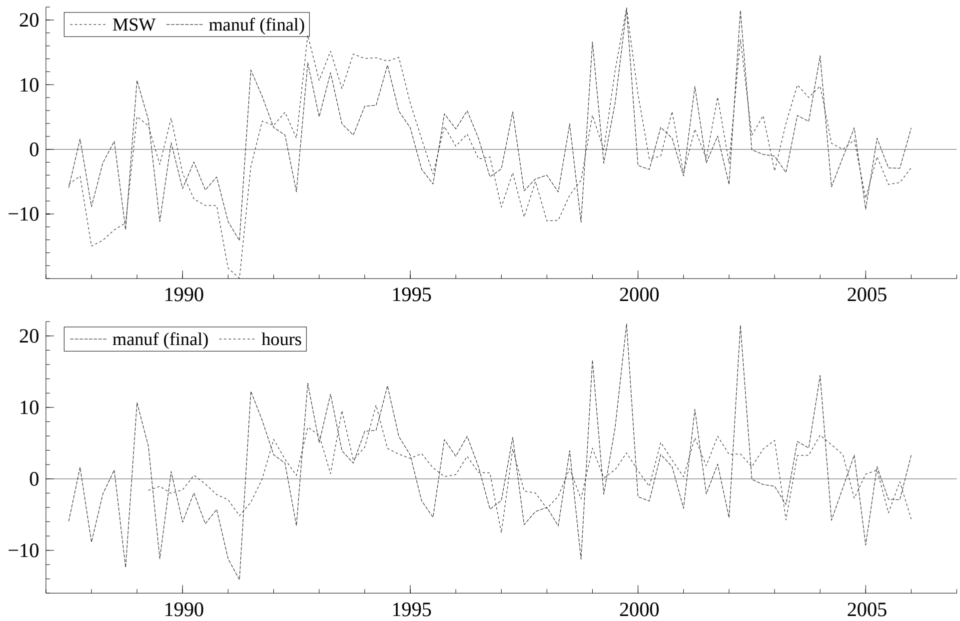
Figure 1 plots the MSW nowcast and RBNZ's hours paid nowcast against the final estimates for manufacturing GDP growth. Data on hours paid are available from 1989Q2 only. Figure 2 plots the corresponding nowcasts for GDP growth with the RBNZ1 estimates from 2000 also shown. Figures 1 and 2 show that the MSW indicator performs well at tracking the official data. It performs particularly well for manufacturing GDP growth relative to the hours paid benchmark.

Table 1 summarises the performance of the indicators. The MSW indicator provides more accurate early estimates of output growth than traditional indicators based on aggregate data. The higher correlation of the MSW indicator shows that a stronger signal about official data may be recovered from the panel data rather than from the aggregate survey data.

The MSW indicator also provides a superior in-sample fit to the RBNZ nowcasts. The RBNZ nowcast for manufacturing GDP growth performs comparably to traditional indicators constructed from survey data. This suggests that, even without exploiting disaggregate survey information, one can construct a useful rival to the RBNZ nowcast just from the aggregate findings from QSBO.

[5] We note, however, there is an argument for considering the annual growth rate as the survey question asks firms to compare their experience in the last three months relative to the same three-month period in the previous year.

Figure 1
In-sample comparison of quarterly manufacturing GDP growth predictions

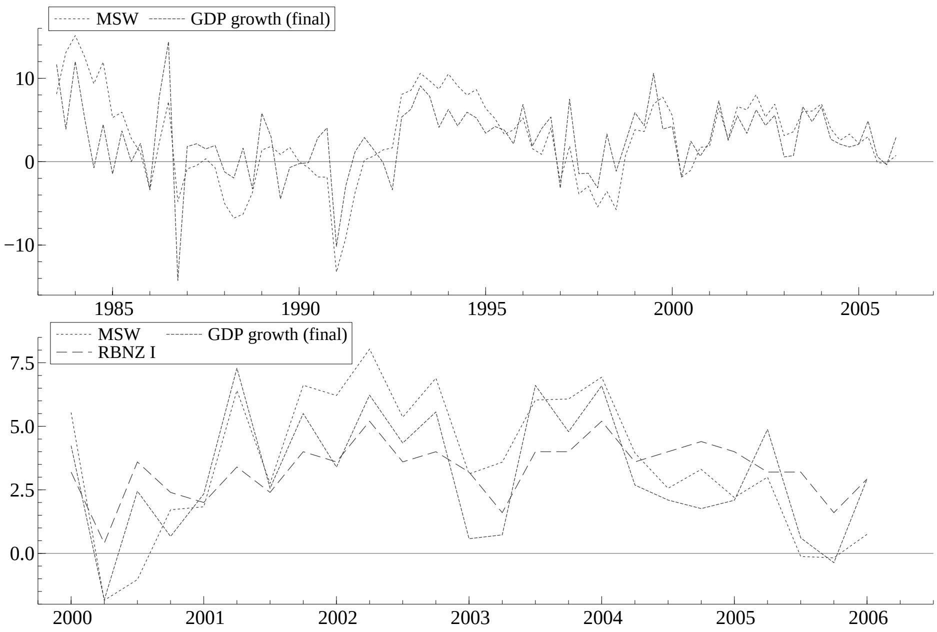


For GDP growth, the RBNZ nowcast (RBNZ1) contains a strong signal about GDP growth with a correlation of 0.71. This can be contrasted with a correlation of only around 0.4 for the aggregate survey-based nowcasts. The MSW indicator, however, while disadvantaged in the sense it is formed ahead of this ‘first-pass’ nowcast, provides an even stronger signal over both the full-sample period and post 2000Q1, where its correlation is 0.83. It is noteworthy that the RBNZ2 indicator (which is in fact a forecast) performs poorly with a correlation of only 0.07. Thus, waiting for the release of QSBO appears to be worthwhile.

6 Out-of-sample forecasting experiment

Given the relatively good in-sample fit of the MSW indicator, we now examine whether this superiority extends out-of-sample. To evaluate how accurate survey-based early estimates of output growth would have been out-of-sample, we conduct recursive experiments designed to mimic the “real-time” application of the different quantification approaches. This involves, for each recursive sample, re-estimation of both the ordered logit models and the moments of $f(.)$.

Figure 2
In-sample comparison of quarterly GDP growth predictions



Exploiting the real-time data triangles complied by the RBNZ, we assess performance against both the first official estimate and the latest (or 'final') estimate.[6] These real-time data triangles also mean that we only use information genuinely available at the time (at each recursive sample) to produce the nowcasts. This means, for example, the firm-level ordered discrete choice models underlying the MSW indicator are estimated using the latest vintage of the official output data in each recursive sample.

The out-of-sample analysis for GDP is conducted over 52 quarters from 1993Q2 to 2006Q1. This involves, for example, using real-time official data for GDP from the 06/07/1993 vintage (which contains GDP data from 1981Q2 to 1993Q1), to nowcast GDP growth in 1993Q2 using the QSBO data published early in 1993Q2. The out-of-sample analysis for manufacturing GDP is conducted over 25 quarters from 2000Q1 to 2006Q1. The shorter out-of-sample period is explained by the shorter in-sample period for manufacturing GDP growth. Official manufacturing data are available from 1987Q2, but real-time vintages are only available from 2000Q1. Real-time nowcasting, therefore, begins using this 2000Q1 vintage to

[6] Values are seldom truly final because data revisions and the arrival of new data continue.

Table 1
In-sample correlations with official data

Correlation	Manufacturing GDP Growth	GDP Growth
	1987Q3-2006Q1	1983Q3-2006Q1
MSW (> 20)	0.75	0.71
MSW (> 20): 2000Q1-	-	0.83
MSW (pooled)	0.46	0.45
RBNZ: Hours Paid	0.49	-
RBNZ1 nowcast: 2000Q1-	-	0.71
RBNZ2 forecast: 2000Q1-	-	0.07
BAL	0.46	0.41
CP	0.45	0.40
PES	0.48	0.42
CSW	0.47	0.42

2000Q1- denotes that the correlation is computed from 2000Q1 to the end of the sample.

nowcast the 2000Q1 value for manufacturing GDP growth.

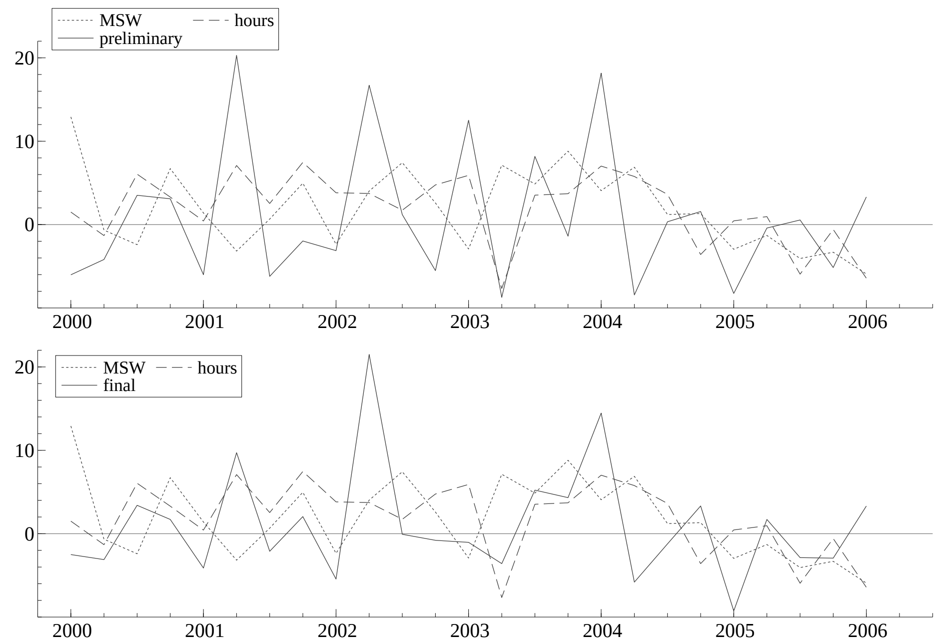
Figures 3 and 4 plot the out-of-sample nowcasts from the MSW indicator alongside the first (or preliminary) and final estimates for the official data. The RBNZ nowcasts are also presented.

It is evident that there is considerable volatility in the official data for manufacturing GDP growth, especially in the first estimates, and it is difficult to discern a general pattern in its evolution over the sample period. For GDP growth, both the MSW indicator and RBNZ1 appear to do reasonably well at tracking the general tendency (the rough hump shape) of the official data. This is particularly true for the final data, which are considerably smoother than the first estimate.

Tables 2 and 3 summarise the results of the forecasting experiment more formally by presenting root mean squared error (RMSE) statistics. The accuracy of the nowcasts is evaluated against both the first and final estimates of the official data. If the first official estimate is an efficient forecast of the final estimate, it should have a lower standard deviation since the forecast should be less variable than the subsequent realisation (which is affected by shocks that occur after the first estimate was made).

Table 2 shows the results for manufacturing GDP growth while tables 3 and 4 show the results for GDP growth over two different out-of-sample periods. The shorter out-of-sample period is provided to facilitate comparison of the MSW in-

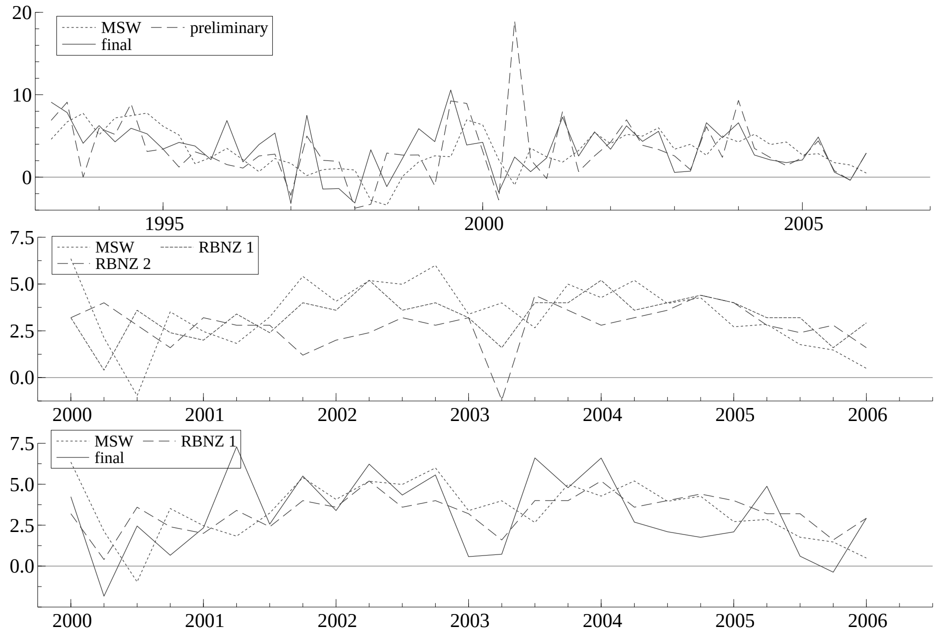
Figure 3
Out-of-sample comparison of quarterly manufacturing GDP growth (final or latest vintage) against MSW and RBNZ (hours) nowcasts



dicator against the RBNZ nowcast (RBNZ1), which is available from 2000 only. On a RMSE basis, there is little to choose among the competing nowcasts, with Diebold-Mariano tests confirming no statistical difference (at 99 per cent confidence) with p-values of zero. The RBNZ nowcast (using data on hours paid) is better than the baseline MSW indicator, but it performs comparably to the aggregate survey-based nowcasts. Table 2 also reveals that the nowcasts, in general, are more closely related to the final rather than the first estimate of the official data. We explore this further in section 6.1.

Table 2 also confirms what was observed in figure 1: manufacturing GDP growth is a noisy series. The nowcasts struggle to have a RMSE lower than the standard deviation computed over the out-of-sample period, so that the nowcasts struggle to beat the (ex post) unconditional mean. We should expect this to remain a good nowcast (as judged by the RMSE alone), assuming we have no reason to expect

Figure 4
Out-of-sample comparison of quarterly GDP growth (final or latest vintage) against MSW and RBNZ (first-pass) nowcasts



a structural break in the unconditional mean. Below, we examine an alternative evaluation criterion to the RMSE.

To determine why the encouraging in-sample performance using the MSW indicator is not translated into an equally good out-of-sample performance, we experimented with two variants of the MSW (baseline) indicator. These indicators depart from being 'pure' real-time indicators by assuming either full-sample information about the prior density (full-sample for prior) or full-sample information about both the prior density and the parameters in the firm-level models (full-sample for prior and logits). The motivation for using full-sample estimates is that if the recursive estimates in the prior density and the parameters in the firm-level models were equal to their full-sample counterparts, the MSW out-of-sample nowcasts would equal the in-sample estimates. These estimates are closely related to the official data and deliver RMSE below those presented in table 2.

Table 2
Out-of-sample performance: RMSE against Final and First quarterly manufacturing GDP growth estimates: 2000Q1-2006Q1

	Final Estimate	First Estimate
Standard deviation	6.61	8.35
Random walk	13.21	15.34
PES	6.64	8.68
CP	6.50	8.44
CSW	6.57	8.60
BAL	6.54	8.50
MSW	7.62	10.13
MSW (Pooled)	10.16	12.31
MSW (Full-sample for prior)	9.98	9.49
MSW (Full-sample for prior and logits)	5.29	8.43
RBNZ: Hours Paid	6.56	7.67

This comparison is particularly informative. It reveals that parameter uncertainty in the firm-level models helps to explain the loss in accuracy of the MSW indicator. When these parameters are not estimated recursively, but set to their full sample values, the RMSE (against the final official estimates) drops from 7.62 to 5.29. This is the lowest RMSE statistic in table 3. It is clearly important, then, to get accurate estimates of these parameter values.

If the relationship between the QSBO responses of firms and official data is stable over time, we should expect more data to improve the accuracy of the MSW indicator as parameter uncertainty diminishes. If, however, there is some instability, this may not be the case. It may then be useful to consider using a rolling rather than a recursive estimation window to avoid considering "out-of-date" data. As we explain below, RMSE and pairwise comparison of the alternative nowcasts is not the complete story. It is possible that a nowcast with a high RMSE still contains valuable information not captured by competing nowcasts; more formally, it may not be 'encompassed'.

Tables 3 and 4 present the results of the out-of-sample simulations for GDP growth. As with manufacturing GDP growth, we find that the (baseline) MSW indicator performs worse than the traditional survey-based indicators. Not surprisingly, perhaps, as it uses more information, RBNZ1 performs best. In contrast, RBNZ2 performs relatively poorly. Again, we find that the survey-based nowcasts deliver a better forecast of the final rather than the first GDP growth estimates. Similarly, as with manufacturing GDP, table 4 shows that it is parameter uncertainty that is

Table 3
Out-of-sample performance: RMSE against Final and First quarterly GDP growth estimates: 1993Q2-2006Q1

	Final Estimate	First Estimate
Standard deviation	3.01	3.83
Random walk	4.97	5.56
PES	2.80	3.69
CP	2.80	3.72
CSW	2.81	3.71
BAL	2.76	3.70
MSW	3.06	4.14
MSW (Pooled)	6.40	7.12
MSW (Full-sample for prior)	2.99	4.09
MSW (Full-sample for prior and logits)	2.48	4.10

adversely affecting the performance of the MSW indicator.

We should not be too surprised by the somewhat disappointing performance of the MSW indicator in these out-of-sample experiments. In an out-of-sample window, itself characterised by considerable volatility/instability, there is no reason to expect a model which explains the historical data well to perform satisfactorily in an out-of-sample basis. (See, for example, Clements and Hendry 1998). Given instabilities, and in our case considerable volatility in the out- of-sample period, robust (that is, simple) forecasting models are known to be hard to beat. Indeed, in our application, the unconditional mean is actually hard to beat. This might help to explain why the simpler, more parsimonious, traditional quantification approaches do quite well relative to MSW. MSW is designed to extract the best in-sample fit between the survey data and the output data. As these output data are by necessity historical, the value of MSW on an out-of-sample basis naturally depends on how informative the past is about the future.

6.1 Predicting revisions using survey data

In this section, we evaluate the information content of the MSW nowcast in terms of its capacity to explain the final estimate of the official data, relative to the other nowcasts and the first estimate. This task amounts to testing whether the survey data, and specifically the MSW indicator, can be used to explain revisions to the official data. Accordingly, we use the nowcasts from the out-of-sample simulations in conjunction with the first official estimate to test the predictability of revi-

Table 4
Out-of-sample performance: RMSE against Final and First quarterly GDP growth estimates: 2000Q1-2006Q1

	Final Estimate	First Estimate
Standard deviation	2.40	4.16
Random walk	5.19	6.72
PES	2.18	4.32
CP	2.09	4.29
CSW	2.10	4.32
BAL	2.06	4.31
MSW	2.39	4.83
RBNZ1	1.76	3.64
RBNZ2	2.57	4.29

sions. Specifically we test for 'news' versus 'noise' using Mincer-Zarnowitz tests along the lines of Faust *et al* (2005). In effect, this lets us determine the weight that should be attached to each nowcast. We can then test whether the weight on a given nowcast is zero. This would imply that this nowcast is encompassed by the others and offers no value-added information relative to the alternatives.

The null hypothesis of 'news' (or efficiency of the first estimate) amounts to testing, via a Wald or F-test that is robust to serial correlation and heteroscedasticity), the joint hypothesis that $b_0 = 0$ and $b_1 = 0$ in the regression: $R_t = b_0 + b_1 P_t + \varepsilon_t$, where R is the revision, defined as the difference between the final and preliminary (P) official estimate. Under the 'noise' model, the preliminary estimate helps predict the subsequent revision R, implying a rejection of the null. The 'news' model, on the other hand, implies that any extraneous information known at the time the preliminary estimate was formed should be orthogonal to R.

Table 5 summarises the results comparing MSW (the baseline case) against the relevant RBNZ alternative. Three sets of results are presented for GDP growth and manufacturing GDP growth: (i) a regression of the revision on the MSW indicator, the RBNZ indicators (RBNZ1, RBNZ2 or hours paid) and the preliminary or first official estimate (denoted *Revision*); (ii) a regression of the final estimate on these three explanatory variables (denoted *Final*) and (iii) a regression of the first estimate on the MSW and RBNZ nowcasts (denoted *First*). The first regression is a Mincer-Zarnowitz test that examines whether there is a predictable or systematic component to revisions that can be exploited. The second regression simply re-expresses this first regression. It takes the form of a traditional forecast combination regression and indicates the optimal weights on the competing

Table 5
Forecast combination and explaining revisions

	c	MSW	RBNZ1	First	R^2	F
GDP growth: 2000Q1-2006Q1						
Revision	-1.95	0.4	0.89	-0.79	0.81	23.47 (0.00)
	(-1.97)	(2.06)	(1.72)	(-4.85)		
Final	-1.95	0.4	0.89	0.21	0.58	29.38 (0.00)
	(-1.97)	(2.06)	(1.72)	(1.28)		
First	-0.99	-1.31	2.75		0.53	20.48 (0.00)
	(-0.55)	(-1.85)	(4.35)			
	c	MSW	RBNZ2	First	R^2	F
Revision	-1.09	0.72	0.15	-0.63	0.78	19.47 (0.00)
	(-0.8)	(3.33)	(0.42)	(-6.06)		
Final	-1.09	0.72	0.15	0.37	0.51	24.56 (0.00)
	(-0.8)	(3.33)	(0.42)	(3.55)		
First	5.73	-0.65	0.03		0.07	6.74 (0.00)
	(1.54)	(-0.83)	(0.06)			
GDP growth: 1993Q2-2006Q1						
Revision	1.03	0.35		-0.61	0.50	16.28 (0.00)
	(2.03)	(2.7)		(-4.4)		
Final	1.03	0.35		0.39	0.39	46.85 (0.00)
	(2.03)	(2.7)		(2.82)		
First	2.4	0.28			0.03	20.61 (0.00)
	(2.18)	(1.14)				
Manufacturing: 2000Q1-2006Q1						
	c	MSW	hours	First	R^2	F
Revision	-0.08	0.38	-0.15	-0.27	0.52	5.6 (0.00)
	(-0.14)	(2.97)	(-1.37)	(-2.13)		
Final	-0.08	0.38	-0.15	0.72	0.76	17.51 (0.00)
	(-0.14)	(2.97)	(-1.37)	(5.61)		
First	0.1	-0.43	0.86		0.21	2.10 (0.13)
	(0.11)	(-1.79)	(2.93)			

HAC robust t-values in parentheses. F-tests are presented with the p-value in parentheses.

nowcasts, letting us test whether any given nowcast is encompassed by the others. The third regression is also a forecast combination regression. It is relevant for the case when the user is interested in predicting the first estimate rather than its revisions.

It is clear from table 5 that the first official estimates are not rational expectations of the final estimates, and that one can explain a statistically significant amount of the variation in the revision (about its mean). The null hypothesis of the 'news' model is rejected with p-values of zero.[7] It is also apparent that in order to construct the best possible indicator it is important to look beyond the preliminary data. In particular, it is the MSW indicator and not the RBNZ nowcast that helps to explain the revisions. This is the case even when MSW is compared with RBNZ1 rather than RBNZ2. Equally, the MSW indicator is statistically significant in the revisions equation for manufacturing GDP growth, unlike the hours paid nowcast. This outcome is encouraging and suggests that the MSW indicator helps to explain revisions to official data over and above the information contained in the first release of the official data and the RBNZ nowcasts.

Finally, we need to compare MSW against traditional survey-based indicators, since alternative means of extracting information from QSBO may perform just as well. It may be that QSBO is useful irrespective of how the firm-level data are quantified and aggregated. We consider the Pesaran (PES) indicator as representative of the traditional aggregate indicators. Table 6 presents the results.

Comparing tables 5 and 6 over the shorter sample period from 2000Q1 to 2006Q1, we find that MSW and PES offer similar explanatory power for revisions. In addition, one indicator can be substituted for the other with little loss. MSW offers a better explanation for the first official estimate than PES. Over the longer out-of-sample period 1993Q2 to 2006Q1, PES appears to offer a little less explanatory power for revisions than MSW, and, when both indicators are considered together, MSW is more significant than PES.

Summarising, it seems fair to say that the MSW indicator offers, at best, marginal gains relative to PES when predicting revisions. Use of QSBO in general, however, is clearly helpful in explaining revisions to GDP growth. It is also worthwhile recalling that we have been considering the MSW baseline indicator assuming parameter instability. It is possible that the performance of MSW will improve as parameter uncertainty diminishes with more data. Table 6, however, indicates a greater gain for MSW relative to PES when attempting to explain revisions to manufacturing GDP growth. The MSW indicator offers a better explanation than

[7] This is consistent with our finding that the standard deviation of the first estimates is higher than the standard deviation of the final estimates (see tables 2 and 3).

Table 6
Forecast combination and explaining revisions

	GDP growth: 2000Q1-2006Q1						
	c	PES	MSW	RBNZ1	First	R^2	F
Revision	-3.27	0.76	-	0.86	-0.80	0.82	24.72 (0.00)
	(-3.76)	(2.25)		(1.73)	(-5.49)		
First	1.58	-1.84		2.68		0.43	15.71 (0.00)
	(0.411)	(-1.28)		(3.43)			
Revision	-3.01	0.58	0.18	0.77	-0.78	0.82	19.17 (0.00)
	(-2.59)	(1.00)	(0.53)	(1.45)	(-5.07)		
	GDP growth: 1993Q2-2006Q1						
Revision	0.71	0.50	-	-	-0.62	0.48	15.20 (0.00)
	(1.14)	(2.37)			(-4.14)		
Revision	0.94	0.06	0.32		-0.61	0.50	11.97 (0.00)
	(1.38)	(0.18)	(1.39)		(-4.31)		
	Manufacturing: 2000Q1-2006Q1						
	c	PES	MSW	hours	First	R^2	F
Revision	0.26	0.54	-	-0.16	-0.29	0.45	4.33 (0.01)
	(0.53)	(1.77)		(-1.08)	(-2.09)		
Revision	-0.06	0.08	0.35	-0.16	-0.27	0.52	4.33 (0.01)
	(-0.12)	(0.23)	(2.46)	(-1.18)	(-2.08)		
First	0.06	-0.26	-0.33	0.90		0.21	1.53 (0.23)
	(0.06)	(-0.26)	(-0.69)	(2.65)			

HAC robust t-values in parentheses. F-tests are presented with the p-value in parentheses.

any of the alternative indicators.

7 Conclusion

Business surveys have long been key sources of information for central banks and policymakers. Traditional methods of converting qualitative survey responses to a quantitative measure have used aggregate survey data, with the net balance statistic being the best known. The regression and probability methods are also well-established, especially in the academic literature. This paper investigates whether an improved real-time signal of official output changes can be derived from firm-level (or panel) survey responses. The paper also considers the important issue of the ability of survey data to anticipate revisions to official output data.

We find, overall, that exploiting the panel dimension to qualitative survey data gives a better in-sample signal about official output data than traditional methods and RBNZ nowcasts. Out-of-sample, it is matters less how survey data are quantified with simpler and more parsimonious methods hard to improve upon. We also find that the MSW panel-based method of quantifying survey data helps to explain revisions to official data.

References

Bates, J M and C W J Granger (1969), "The combination of forecasts," *Operational Research Quarterly*, 20, 451–468.

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

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

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 I Begg and S G B Henry, Cambridge University Press, Cambridge.

Driver, C and G Urga (2004), "Transforming qualitative survey data: performance comparisons for the UK," *Oxford Bulletin of Economics and Statistic*, 66, 71–89.

Faust, J, J Rogers, and J Wright (2005), "News and noise in G-7 GDP announcements," *Journal of Money, Credit, and Banking*, 37, 403–417.

Granger, C W J and R Ramanathan (1984), "Improved methods of combining forecasts," *Journal of Forecasting*, 3, 197–204.

Mitchell, J, R Smith, and M Weale (2005a), "Aggregate versus disaggregate survey-based indicators of economic activity," *National Institute of Economic and Social Research*, *Discussion Paper*, 194, (Revised).

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

Mitchell, J, R Smith, and M Weale (2006), "A Bayesian indicator of manufacturing output from qualitative business panel survey data," *National Institute of Economic and Social Research*, *Discussion Paper*, 261.

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

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

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

Pesaran, M H and M Weale (2006), "Survey expectations," in *Handbook of Economic Forecasting*, eds G Elliott, C W J Granger, and A Timmermann, 715–776, North-Holland, Amsterdam.

Sleeman, C (2006), "Analysis of revisions to quarterly GDP – a real-time database," *Reserve Bank of New Zealand Bulletin*, 69(1), 31–44.

Smith, J and M McAleer (1995), "Alternative procedures for converting qualitative response data to quantitative expectations: An application to

Australian manufacturing," *Applied Econometrics*, 10, 165–185.
Stark, T (2002), "A summary of the conference on real-time data analysis," *Federal Reserve Bank of Philadelphia Business Review*, 5–11.