Ask Me No Questions and I'll Tell You No Lies: Comparison of Qualitative and Quantitative Firm Performance Measures

Richard Fabling[*], Arthur Grimes[†] and Philip Stevens[‡]

May 2007

DRAFT: NOT FOR QUOTATION OR CITATION WITHOUT THE AUTHORS' PERMISSION

Abstract

Many analyses of firm performance are based upon self-reported measures. However, not only are these likely to be more subject to general reporting error than alternative official sources, but also measures of relative performance may be subject to the biases observed in the psychology literature. One of these is that most people think they are above average. In this paper we shall consider both absolute and relative performance measures reported in the Business Operations Survey with alternative measures taken from official sources, brought together in the Longitudinal Firm Performance Dataset. We also consider the implications of using any such dissonance as a measure of one aspect of managerial quality: the ability to understand the business environment in which the firm operates.

JEL Classifications: C80; C81; D24; L25

Keywords: Micro data; subjective data: firm performance; labour productivity; management capability

[*] Ministry of Economic Development (MED); seconded to Statistics New Zealand; Affiliate – Motu Economic and Public Policy Research (Motu)
[†] Motu; University of Waikato; contracted to MED for this research
[‡] MED; Visiting Fellow – National Institute of Economic and Social Research, London

Disclaimer

Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The research was funded by the Ministry of Economic Development and supported by Statistics New Zealand as part of the Improved Business Understanding via Longitudinal Database Development project (IBULDD). Any views expressed are those of the authors and do not purport to represent Statistics New Zealand, the Ministry of Economic Development or any other agencies to which the authors are affiliated.

The results of this study are based in part on tax data supplied by Inland Revenue to Statistics New Zealand under the Tax Administration Act. This tax data must be used only for statistical purposes, and no individual information is provided back to Inland Revenue for administrative or regulatory purposes. Any discussion of data limitations or weaknesses is in the context of using the data for statistical purposes, and is not related to the ability of the data to support Inland Revenue's core operational requirements. Careful consideration has been given to the privacy, security and confidentiality issues associated with using tax data in this project. In particular, in the IBULDD dataset, individuals' tax data has been aggregated to the firm-level. Furthermore, only people authorised by the Statistics Act 1975 are allowed to see data about a particular firm.

1 Introduction

Many analyses of firm performance are based upon self-reported measures. However, these are likely to be subject to reporting error and/or perception biases. Self-reported measures of firm performance are likely to be more subject to general reporting error than alternative official sources. Respondents are often under legal obligations to provide correct data for official purposes such as tax reporting. Even when the return of questionnaires duly filled in is a compulsory requirement, the incentives and thus the time taken to fill in surveys are lower (particularly if one member of staff does not hold all of the required information). In addition to this, self-reported measures of *relative* performance may be subject to the biases observed in the psychology literature. One of these is that most people think they are above average.

In this paper we shall consider both absolute and relative performance measures reported in the Business Operations Survey (BOS) with alternative measures taken from official sources, brought together in the Longitudinal Firm Performance Dataset (LFPD). We also consider the implications of using any such dissonance as a measure of one aspect of managerial quality: the ability to understand the business environment in which the firm operates. At the heart of our analysis is the data upon which we base our comparisons, these are described in section 3. We describe our results in section 4. In section 5 we consider the implications of using the dissonance between objective and subjective measures of performance as a measure of managers' ability to understand the business environment in which the firm operates. Section 6 concludes.

2 Background

Many analyses of firm performance are based upon self-reported measures. As we have noted above, self-reported measures of firm performance are likely to be subject to reporting error and/or perception biases. The first of these is likely to be because, unlike surveys, respondents providing data for official purposes, such as tax reporting, are usually under legal obligations to do so correctly. It is likely that the incentives and thus the time taken to fill in surveys are lower than for official purposes. In a survey, the respondent may not hold all of the required information and is required to either discuss with staff that do, or to estimate it themselves. Each link makes the chain weaker.

Another problem with survey information is that much of it is subjective, rather than objective. Examples of subjective information include questions on job and life satisfaction, or

assessments of the business environment by firms. These questions may be subject to cognitive problems (e.g. related to the ordering or framing of questions), social desirability issues ('what do you want me to say?', 'what do I want you to hear me say?') and/or situations in which objective answers do not exist or for which people *cannot* make the relevant choices (Bertrand and Mullainathan, 2001). Previous work investigating how managers respond to surveys show that respondents reply differently to subjective and objective questions (Hillage *et al.*, 2002; Mason, 2005; Forth and McNabb, 2006a,b). This has important ramifications regarding how such measures are interpreted.

Are We All Better Than Average, On Average?

One potential bias that is likely to affect self-reported measures of *relative* performance is the tendency to for people to believe (or at least report) that they are above average. According to the survey of the psychology literature by Taylor and Brown (1988) people have unrealistically positive views of the self. For example, evidence suggests that managers are inclined to believe that they are superior to the average manager (Larwood and Whittaker, 1977), and entrepreneurs perceive their own chance for success as being higher than that of their peers (Cooper, Woo and Dunkelberg, 1988). The issue of overconfidence is considered an important one in investment theory, although here overconfidence is often modelled as underestimation of the variance of signals (De Bondt and Thaler, 1995).

3 Data

The Business Operations Survey (BOS) data are matched to data obtained from Statistics New Zealand's Longitudinal Firm Performance Dataset (LFPD). The LFPD is built around the Longitudinal Business Frame (LBF) to which are attached, among other things, Goods and Services Tax (GST), financial accounts (IR10s) and aggregated Pay-As-You-Earn (PAYE) returns all provided by the Inland Revenue Department (IRD). The full LFPD is described in more detail in Fabling *et al.* (2007).

The survey data considered in this paper relate to the Business Operations Survey (BOS) 2005. The BOS is an annual three part modular survey, which began in 2005. The first module is focussed on firm characteristics and performance. The second module alternates between biennial modules relating to innovation and business use of ICT. The third module is a contestable module that enables specific policy-relevant data to be colleted on an *ad hoc* basis. BOS is conducted using two-way stratified sampling, with stratification on rolling-mean

employment (RME) and two-digit industry according to the ANZSIC system[1]. The 2005 survey was sent to 6,979 enterprises with a total of 5,595 usable responses returned (a response rate of 80.2% after adjusting for ceases)[2]. It is important to note that in common with many surveys conducted by Statistics New Zealand (SNZ) the survey is statutory and the front page bares the imprimatur 'The taking of this survey has been approved by the Minister of Statistics and the return of this questionnaire, duly filled in and signed, is a compulsory requirement under the Statistics Act 1975'. Because of this, BOS has a considerably higher response rate than comparable surveys internationally. The implications for data quality are uncertain. Whether this requirement increase the quality of responses or simply brings into the sample a number of firms who will spend less time and effort on the survey remains to be seen[3].

The administrative data to which we shall be comparing to BOS have three sources, counts of employees from PAYE returns, the Business Activity Indicator (BAI) dataset and IR10 forms. The BAI is derived from GST data, with the main manipulations applied being temporal and group return apportionment and limited imputation for single missing returns. In this paper, the BAI is the source for data on sales of goods and services, and purchases. Financial accounts returns (IR10) are the source for information on purchases, profits, opening and closing stock. We also use IR10 sales for comparative purposes, with the difference between two alternative administrative (tax) sources providing some context for our comparison with BOS. The variables used in this paper are discussed in more detail in the Appendix.

In order to make the appropriate comparisons, it is important to ensure that information relates to the same financial year. Respondents to the BOS are asked to state 'the balance date of the financial accounts which you will use for this questionnaire' (Section A, Part I, Question 6). This date is used to match PAYE, IR10 and BAI data to the appropriate financial year.

4 Results

Our comparisons take three forms. First, we compare the reported financial data in BOS with that BAI and IR10 data. Next, we compare self-reported subjective measures of performance with self-reported 'objective' financial measures. Finally, we compare respondents' perception

[1] Note that there was some minor additional stratification conducted at the three-digit level.
[2] There was some over-sampling of BOS 2005. However, because over-sampled unit responses were not subject to the same quality assessment by Statistics New Zealand, we have excluded them from our current analysis. These additional data are available for future analysis.
[3] It may be possible to gain a greater understanding of these data quality issues using information on follow up requests made by Statistics New Zealand to firms that had not yet returned their completed surveys and the reported time taken to fill out the survey (Question 39 'How long did it take to complete this questionnaire?'). We leave this for future work.

of changes in key variables with changes from the administrative data. In what follows, in order to comply with Statistics New Zealand confidentiality requirements, we have rounded unweighted counts of firms to the nearest five. All other figures are unmodified.

4.1 Comparing Self-reported and Official Quantitative Data

We begin our analysis with a comparison of financial data reported in the first section of the BOS with that available from alternative sources. Surveys are often used as a source of financial information and are used as a performance metric against which to evaluate determinants of interest, e.g. innovation, business practices, exporting behaviour (e.g. Fabling, 2007a; Fabling and Grimes, forthcoming).

Sales

Our first comparison is sales of goods and services. Respondents to the BOS are asked to supply GST exclusive amounts when supplying financial information. Where respondents have indicated that the figures do in fact include GST, GST exclusive figures are computed by multiplying by 8/9. This only affects a very small proportion of observations. The data in the BAI are GST inclusive and those in the IR10s GST exclusive. Because of this, in what follows we also adjust the figures taken from BAI returns[4].

Table 1 and Table 2 show the results of such a comparison. We can see from the first set of columns in Table 1 that we are less likely to have sales figures from IR10 returns and that firms which do not return IR10 forms appear to be larger than average (for our sub-sample of firms taking part in the BOS).[5] Because of this, and analysis that suggests that BAI purchases data are more appropriate than IR10 purchases data for constructing productivity measures (Cox, 2007), we concentrate on BAI sales data from here on in[6].

The second set of columns provide data on sales for firms for whom we have all sets of data. The figures in the BOS are slightly higher than the figures for which we have BAI sales figures, which are in turn slightly higher than those for which we have IR10s. However, these differences are very small – indeed that between the highest and lowest means is less than 5%. The final columns test the hypothesis that the IR10 and BAI figures are insignificantly different from those in the BOS. We cannot reject the hypothesis of equality between the two alternative

[4] For more details, see the Appendix

[5] Administrative data could be missing because (i) the firm's primary business consists of GST exempt financial activities, in industries K and L; (ii) firm non-response; (iii) the firm was created or destroyed during its filing period; (iv) the firm appears to exist, but in fact does not (a GST return may have been filed when the firm was being wound down, for example).

[6] We need sales and purchases from an internally consistent source for the calculation of value-added.

official sources of sales data and that obtained from the BOS[7]. When we restrict the comparison to BAI and BOS, the number of observations on which we can base our comparison increases considerably. The *F*-test of equality also increases (i.e., the test that they are significantly different is even less significant than for the smaller subgroup for which we have all three sets of data).

The correlations between the three measures of sales can be seen in Table 2. These are pair-wise correlations and so are based on all firms for which we have data on the two respective sales measures. There is a high degree of correlation between the three measures of sales.

Table 1 Comparing Sales from Alternative Sources

	Separate Samples			Common Sample		Test of equality with BOS	
	Mean	*Std. Err.*	*Obs.*	*Mean*	*Std. Err.*	*F*	*p*
				Common sample = 3,960 observations			
IR10	3,854,707	153,064	4,010	3,891,592	155,602	1.68	0.1947
BAI	6,889,728	328,751	5,385	3,940,389	154,207	0.95	0.3296
BOS	7,288,049	376,577	5,525	4,070,508	197,747	.	.
				Common sample = 5,325 observations			
BAI	.	.	.	6,967,994	334,646	0.02	0.8878
BOS	.	.	.	6,994,962	365,861	.	.

- *Figures based in sample strata and weights (except the observations which relate to unweighted data)*
- *All figures exclude GST.*
- *Observations rounded to nearest five for confidentiality reasons*
- *Figures for GST inclusive BAI sales are brought into line with GST-exclusive BOS figures by multiplying non-zero-rated GST sales by 8/9.*

[7] Note that these comparisons are for the subset of firms for which we have all three measures.

Table 2 Correlations between Measures of Sales

		Pearson						*Spearman (rank)*		
		Unweighted			*Weighted*					
		IR10	BAI	BOS	IR10	BAI	BOS	IR10	BAI	BOS
IR10	ρ	1			1			1		
	p	.			.			.		
	Obs	4,010			.	.	.	4,010		
BAI	ρ	0.8909	1		0.8851	1		0.8972	1	
	p	(0.000)	.		(0.000)	.		(0.000)	.	
	Obs	4,010	5,385		.	.	.	4,010	5,385	
BOS	ρ	0.8793	0.8550	1	0.8294	0.8574	1	0.8854	0.9034	1
	p	(0.000)	(0.000)	.	(0.000)	(0.000)	.	(0.000)	(0.000)	.
	Obs	3,960	5,325	5,525	.	.	.	3,960	5,325	5,525

Profits

The term profit can be interpreted in a number of different ways. Our ability to make comparisons is subject to data availability. GST-based data does not contain sufficient information to allow a comparison with BOS. For a comparison of administrative and survey measures of profitability, we use data from IR10 returns. One disadvantage of this is that it reduces our sample size because of the lower response/submission rates for IR10 returns. The administrative source of profits is therefore total taxable profit from the IR10 return. We calculate operating profit from the data in the BOS as total operating revenue less operating expenses[8].

It is clear from Table 3 that total taxable profits from firms' IR10 returns and operating profits from BOS are measuring somewhat different quantities. Taxable profits from the IR10s are considerably smaller than operating profits calculated from the BOS. This is not due to the larger average size of IR10 non-respondents (as we saw from the sales figures in Table 1), although the figures are marginally closer when we consider the common sample. Even in the sample of firms for which we have both sources of profits, those calculated from BOS are on average five times larger than the sum firms report as taxable profits[9].

[8] See the data appendix for the definitions of operating expenditure and revenue in the BOS.
[9] The Wald-test easily rejects the hypothesis of equality.

Table 3 Comparing Profits from Alternative Sources

	Separate Samples			Common Sample		Test of equality with BOS	
	Mean	*Std. Err.*	*Obs.*	*Mean*	*Std. Err.*	*F*	*p*
				Common sample = 3,945 observations			
IR10 Taxable Profit	234,996	14,186	4,010	239,988	14,570	15.55	0.001
BOS Operating profit	1,828,718	245,568	5,500	1,177,714	238,440	.	.

- *Figures based in sample strata and weights (except the observations which relate to unweighted data)*
- *All figures exclude GST.*
- *Figures for GST inclusive BAI sales are brought into line with GST-exclusive BOS figures by multiplying non-zero-rated GST sales by 8/9.*

One explanation for this is that the two concepts *should* be measuring different things, because of different definitions. Another is that there are clearly incentives to reduce the amount of profits liable to taxation. This may have a temporal dimension. For example, firms may shift profits across years for tax or other purposes and we have just chosen a year when firms have tended to over-report expenditure. However, given that respondents probably had access to the same accounts that were used for the IR10s, this explanation seems unlikely. This is an issue for further analysis, which we intend to consider in the future.

Despite the difference in levels of profits they are significantly correlated. In particular, note the much higher rank correlation. This is consistent with the hypothesis that one profit definition is a monotonic, non-linear transformation of the other. Thus, analysis using these different measures of profits might come to similar conclusions, but this is by no means certain, especially where the raw data (rather than rankings) are used.

Table 4 Correlations between IR10 and BOS Measures of Profits

	Pearson		*Spearman (rank)*
	Unweighted	Weighted	Unweighted
ρ	0.2767	0.1571	0.6739
p	0	0	0
Obs	3,945	.	3,945

Employment

Employment raises rather different issues to the financial variables. Employees can be full or part-time; they can be temporary or permanent; they can be employed for the whole of the year,

or part of it; some staff may not be employees (eg, working proprietors). In this paper our administrative measure of employment is defined as an average of the twelve monthly PAYE employee counts in the year (known as rolling mean employment, or RME). Because this measure excludes working proprietors we add one to the RME of every firm-year observation.

The comparison of RME employment and that reported in BOS are presented in Table 5. The mean total employment reported in the BOS is very similar to that of (employer-adjusted) RME from the PAYE data, although the variance is higher. The results of the F-test suggest that the series are statistically indistinguishable. This result is mirrored in the extremely high degree of correlation reported in Table 6.[10]

Table 5 Comparison of PAYE-based RME and Self-reported Employment in BOS

	Separate Samples			Common Sample		Test of equality with RME	
	Mean	*Std. Err.*	*Obs.*	*Mean*	*Std. Err.*	*F*	*p*
				Common sample = 5,350 observations			
RME	31.30	1.195	5,350	31.30	1.195		
BOS Working proprietors	1.86	0.077	5,595	1.87	0.079		
BOS Employees	29.39	4.096	5,595	29.17	4.223	0.27	0.6035
BOS Total Employment	31.25	4.098	5,595	31.04	4.225	0.00	0.9487

- *Observations rounded to nearest five for confidentiality reasons*
- *Weighted and stratified*

Table 6 Correlations between Measures of Employment

		Pearson			*Spearman (rank)*		
		RME	BOS Employees	BOS Total Emp	RME	BOS Employees	BOS Total Emp
RME	ρ	1			1		
	p	.			.		
	Obs	5,350			5,350		
BOS Employees	ρ	0.3899	1		0.8997	1	
	p	0	.		0	.	
	Obs	5,350	5,595		5,350	5,595	
BOS Total Employment	ρ	0.3899	0.9998	1	0.9058	0.9872	1
	p	0	0	.	0	0	.
	Obs	5,350	5,595	5,595	5,350	5,595	5,595

- *Observations rounded to nearest five for confidentiality reasons*

[10] At first glance, one might conclude from Table 5 that the assumption of one working proprietor per firm is an understatement, although this depends on what proportion of working proprietors reported in BOS draw a wage.

Productivity

Productivity is a rather difficult beast to get a handle on. There are a number of ways in which respondents could be answering questions with respect to productivity. First, there are issues about the appropriate denominator, i.e. whether firms are considering labour or multi-factor productivity. Second, there is the issue of the appropriate numerator, i.e. sales, value-added. Not all respondents are likely to have economics degrees and so this question is open to misinterpretation. Earlier evidence (Forth and McNabb, 2006b) suggests that managers are worse at reporting subjective measures of productivity than they are of profitability. Forth and McNabb suggest that this reflects greater clarity as to what is being asked about profitability than productivity (p. 13). In this paper we shall restrict our analysis to the log of labour productivity,[11] equal to the log of value-added less the log of employment. We construct a measure of labour productivity from both administrative data,[12] and solely from within BOS.[13]

As we can see from Table 7, the results from our component analysis carry over into the productivity measure itself. The level of productivity is higher when calculated from the BOS than from the BAI (we can statistically reject the hypothesis that productivity levels are the same). The two series are nevertheless significantly positively correlated, with the degree of correlation between the productivity levels being similar for both types of correlation (most likely because of the log transformation).

Table 7 Comparison of Official and Self-reported Productivity in BOS

			Test of equality		Correlations		
	Mean	*Std. Err.*				*Pearson*	*Spearman*
BAI	10.1273	0.0369	*F*	519.63	ρ	0.4040	0.4986
BOS	11.0346	0.0268	*p*	(0.0000)	*p*	(0.0000)	(0.0000)

- *Weighted and stratified (except Spearman correlations)*

4.2 Comparing Self-reported Subjective/Qualitative and Quantitative Data

In this section we focus on information contained in the BOS itself. In a manner similar to Forth and McNabb (2006b) we compare self-reported qualitative, subjective measures of performance with 'objective' measures derived from financial information reported in the

[11] We use log productivity for comparability with other work. Note that we obtain similar results from calculating labour productivity using the ratio of the levels of the two variables and these results will be presented in a future version of this paper.
[12] Using sales of goods and services minus purchases from the BAI as value-added, and (employer-adjusted) RME.
[13] Using sales of goods and services minus total other operating expenditure from BOS as value-added, and BOS total employment.

BOS[14]. The quantitative financial information is reported in the first part of the BOS, 'Part i: Financial information'. The qualitative performance is contained in the third part of the BOS, 'Part iii: Business performance'. According to the instructions contained in the survey: Part i should be completed by the finance department or the accountant. If the firm does not have an accountant on-site, then firms are instructed that Part I should be completed by the General Manager[15]. The instructions state that Part iii should be completed by the General Manager. We do not have information on who completed each section of the survey. This creates the possibility that there may be some kind of reporting bias introduced. However, with this caveat in mind, the instructions clearly state that the quantitative financial information should be completed either by someone whom has responsibility for finance or by a general manager with reference to an accountant and so we feel reasonably confident that such information is as objective as is possible in such a survey.

The first two subjective measures of performance we consider are relative profitability and productivity. Question 38 of the BOS asks firms how they think their business compares with its major competitors against both these metrics. Taken at face value, Table 8 supports our suspicion that respondents tend on average to consider themselves above average. Approximately twice as many firms feel they are more profitable than their competitors than less, and around six times as many feel they are more productive than less.[16] We should however, bear in mind that this apparent bias could be mitigated by the large number of 'don't know' responses (Fabling and Grimes, forthcoming, find striking similarities between the 'don't know' group and a combined 'low/average' group). For example, firms that reported that they did not know might be less profitable or productive, and either the factors that prevent them knowing are correlated with their low competitiveness (see section 5), or they know the truth and cannot bring themselves to admit it in writing.

[14] In a later version we shall include the comparison of objective data from the official sources with the subjective measures from BOS. As the previous section would suggest, the results from these are similar, but time constraints mean that we have been able to include these in this version the paper.

[15] Although firms are reminded that they 'may need to contact the Accountant to complete some of these questions'

[16] This is consistent with the results for output per head and value-added per head in the Workplace Employment Relations Survey (WERS) reported in Forth and McNabb (2006b), although the respondents to the WERS had five choices ('A lot below average', 'Below average', 'About average', 'Above average' and 'A lot above average'). The majority of (weighted) responses (82%) in Forth and McNabb are either 'about average' (42%) or 'above average' (40%).

Table 8 Self-reported Relative Profitability and Productivity (%)

	Profitability		*Productivity*	
	Unweighted	Weighted	Unweighted	Weighted
Lower than competitors	10.78	9.72	4.95	4.06
On a par with competitors	42.95	45.40	46.24	46.76
Higher than competitors	22.43	17.64	26.97	24.31
Don't know	23.84	27.24	21.84	24.87
	100	100	100	100

- *Table refers to BOS Question 38: 'How do you think this business compares to its major competitors on each of the following? Profitability; Productivity'*

Considering productivity, we can compare the subjective estimate of productivity with a labour productivity measure from the financial information from Part i. The mean values of labour productivity for each of the groups, their standard errors and 95% confidence intervals are presented in Table 9. Those who report that their productivity is lower than competitors clearly have the lowest labour productivity and those who believe they are more productive than their competitors indeed appear to be more productive than those who believe they are on a par with competitors.

Table 9 Subjective and Objective Self-reported Measures of Productivity

	Labour Productivity			
	Mean	Linearised Std. Err.	[95% Conf.	Interval]
Lower than competitors	10.724	0.080	10.567	10.882
On a par with competitors	11.042	0.040	10.963	11.121
Higher than competitors	11.186	0.047	11.094	11.277
Don't know	10.939	0.058	10.826	11.052

- *Table shows values of the log of labour productivity by subjective relative productivity*
- *Weighted and stratified*

We can make this comparison more rigorous by performing Wald tests of the equality of these estimates. These are presented in Table 10. As we can see from the first section of the tables, we can distinguish the groups both from each other and the 'don't know' category. The sole exception is the comparison of the productivity of the firms in the 'on a par with competitors' group, which we cannot distinguish from that of those in the 'don't know' category. This result is confirmed when we perform an OLS regression of labour productivity on dummy variables representing the subjective relative productivity groups and the don't know category (with 'On a par with competitors' as the baseline group) (Table 11).

Table 10 Wald Tests of Equality

		On a par with competitors	Higher than competitors	Don't know (4)
Lower than competitors (1)	*F*	12.51	24.71	4.72
	p	(0.0004)	(0.0000)	(0.0299)
On a par with competitors (2)	*F*		5.42	2.13
	p		(0.0199)	(0.1444)
Higher than competitors (3)	*F*			11.02
	p			(0.0009)
Joint Wald tests		1=2=3	1=2=3=4	
	F	12.46	9.39	
	p	(0.0000)	(0.0000)	

- *Top section of table reports two-way Wald F-test of equality of means of productivity between groups along with probability that the difference is not significantly different from zero.*
- *Bottom section joint test that three categories of relative productivity are equal and that these are jointly equal to the don't know category, respectively.*
- *Weighted and stratified*

We conclude that the subjective data do have predictive power. Note that in Table 9 and Table 11 the 'higher' group has a lower standard error than the 'lower' group and so information on high productivity firms appears more reliable than for low productivity firms, as one might expect intuitively. However, it is important to bear in mind that this test does not take into account the choice of the comparison group. Whilst we consider this briefly when we discuss the changes in firms' market share below, we shall consider the issue of the appropriate reference group firms use when making these comparative statements in more detail in future work.

Table 11 Regression of Labour Productivity on Subjective Measure

Observations	5,175		*F*	9.39
Population	34,760.09		*p*	(0.0000)
Design d.f.	5,175		R^2	0.012
	Coef.	s.e.	*t*	*P > t*
Constant	11.0420	0.0404	273.57	0.0000
Lower	-0.3176	0.0898	3.54	0.0000
Higher	0.1437	0.0617	2.33	0.0200
Don't know	-0.1029	0.0705	1.46	0.1440

- *Table reports results of a linear regression where the dependent variable is the log of labour productivity. Independent variables are dummy variables representing the subjective relative productivity groups and the don't know category (with 'On a par with competitors' as the baseline group)*
- *Observations rounded to nearest five for confidentiality reasons*
- *Weighted and stratified*

The picture is rather murkier when we consider profitability (Table 12 and Table 13), where profitability is defined as profits over sales. Again profitability rises over the groups of firms in each of the self-reported profitability groups, with those in the 'lower than competitors' group tending to have negative profitability. We can distinguish this group statistically from the firms in the 'on a par…' and 'higher than competitors' groups, but not from the 'don't know' group. We also cannot distinguish the 'higher' from the 'on a par' group. However, we cannot do so with quite the degree of certainty than we can with productivity. Again, we obtain similar results when we consider this in the multivariate setting (Table 14).

Table 12 Subjective and Objective Self-reported Measures of Profitability

	Profitability			
	Mean	Linearised Std. Err.	[95% Conf.	Interval]
Lower than competitors	-0.3477	0.2797	-0.8961	0.2006
On a par with competitors	0.1965	0.0254	0.1466	0.2463
Higher than competitors	0.2527	0.0366	0.1809	0.3244
Don't know	-0.6717	0.7660	-2.1734	0.8299

- *Table shows values of profitability by subjective relative profitability*
- *Weighted and stratified*

Table 13 Wald Tests of Equality

		On a par with competitors	Higher than competitors	Don't know (4)
Lower than competitors (1)	*F*	3.75	4.53	0.16
	p	(0.0527)	(0.0334)	(0.9612)
On a par with competitors (2)	*F*		1.59	1.28
	p		(0.2074)	(0.2573)
Higher than competitors (3)	*F*			1.45
	p			(0.2281)
Joint Wald tests		1=2=3	1=2=3=4	
	F	2.81	2.31	
	p	(0.0606)	(0.0740)	

- *Top section of table reports two-way Wald F-test of equality of means of productivity between groups along with probability that the difference is not significantly different from zero.*
- *Bottom section joint test that three categories of relative productivity are equal and that these are jointly equal to the don't know category, respectively.*
- *Weighted and stratified*

Table 14 Regression of Profitability on Subjective Measure

Observations	5,390		F^{3}_{5390}	2.31
Population	34,760		p	0.074
			R^2	0.001
	Coef.	s.e.	t	$P > t$
Constant	0.196	0.025	7.84	0
Lower	-0.544	0.336	-1.62	0.053
Higher	0.056	0.069	0.82	0.207
Don't know	-0.868	1.330	-0.65	0.257

- *Table reports results of a linear regression where the dependent variable is profitability. Independent variables are dummy variables representing the subjective relative profitability groups and the don't know category (with 'On a par with competitors' as the baseline group)*
- *Observations rounded to nearest five for confidentiality reasons*
- *Weighted and stratified*

4.3 Perceptions of Changes in Performance

Up until now we have considered BOS responses to questions about the current financial year. Respondents are also asked about how they believe their businesses performance changed over the last financial year (in regard to their sales, profitability, productivity or market share). In this section we consider firms' perceptions of such changes with alternative objective measures. Using the data from the BAI and IR10s, we can analyse the changes in these key variables in each of these groups.

The only variable that we have not already considered is market share. In order to calculate the change in market share we calculate two market share variables, differing only in their definition of 'market'. In these we define market share as the share of total (BAI) industry sales at the 3-digit and 4-digit level (ANZSIC). There are a number of reasons why one would suspect that it is difficult to define the market within which firms operate. It may be the case that some ANZSIC codes do not necessarily correspond to definitions of a 'market' and this may vary not only by product type, but also by firm. Previous work suggests that there is considerable heterogeneity in how firms respond to questions where they are asked to compare themselves to their competitors (e.g. Mason, 2005). Part of this is due to the nature of the products or services firms are offering, part is due to the fact that firms often only compare themselves to firms with which they feel they *can* compete. Firms producing a much higher quality product are frequently considered to be in a different market.

The results of our comparisons are presented in Table 15[17]. Overall, whilst there is some correlation between the qualitative and quantitative measures for sales, profitability and productivity growth, the results are not very strong. The mean value of sales growth increases across the subjective groups. However, because of the high within-group variance in sales, we cannot reject the hypothesis that the mean value of the change in sales in the 'decreased' and 'stayed the same' groups are the same. However, we can say that those who felt that their sales increased over the last financial year report significantly higher changes in sales than those who felt their sales stayed the same (Wald test = 11.35, $p = 0.000$).

Table 15 Comparing subjective & objective measures of change

	Decreased	*Stayed the same*	*Increased*	*Don't know*	*Joint tests of significance*	
					1=2=3	*1=2=3=4*
Sales	-81,914 (193,002)	289,046 (203,268)	1,061,224*** (105,963)	409,104 (234,503)	16.05 (0.0000)	11.18 (0.0000)
Profitability	-0.068 (0.063)	0.006 (0.022)	1.309 (1.218)	0.036 (0.064)	0.180 (0.3072)	0.89 (0.4444)
Productivity	-0.1853 (0.0933)	-0.0163 (0.0586)	0.1005* (0.0287)	-0.0167 (0.1016)	5.30 (0.0050)	3.71 (0.0111)
Market share (3-digit)	-0.0007 (0.0005)	0.0000 (0.0001)	0.0002* (0.0001)	0.0001 (0.0001)	3.23 (0.0397)	2.56 (0.0532)
Market share (4-digit)	-0.0013*** (0.0006)	0.0003 (0.0002)	0.0005 (0.0001)	0.0003 (0.0002)	4.28 (0.0139)	2.95 (0.0314)

- *Figures relate to absolute change*
- *Figure in parenthesis = Linearised standard error of mean in columns 1-4, p-value of Wald test in columns 5 & 6.*
- *Asterisks relate to results of F-test of difference of means of group with "Stayed the same" groups, *** significant at 1% level, ** significant at 5% level, * significant at 10% level*
- *Weighted and stratified*

The change in profitability is negative in the group of firms who state that profitability has 'decreased' and the increase is larger in those who said that it has increased compared to those who think it stayed the same. Thus, although changes in profitability do rise with self-assessed measures, we cannot differentiate statistically between the change in profitability across the three groups.

[17] Note that because of the issues outlined above, we report comparisons using both the whole sample of firms for whom we have both subjective and official measures of profitability and that excluding the bottom five percent.

The figures for productivity are similar. The mean values of labour productivity is increasing across the subjective change groups, but we can only differentiate the 'increased' and the 'stayed the same' groups statistically, and then only at the 10% level[18].

If we calculate the firms' market share at the three-digit and four-digit ANZSIC level (using the whole population of economically significant firms as the basis for market size), we find low but positive levels correlation between subjective and objective measures of change. Certainly, the firms who felt their market share declined do on average experience negative changes in market share when the market is considered to be either the three- and four-digit industry. However, this is only significantly lower than the 'stayed the same' group for the four-digit measure. Given that we cannot distinguish the 'increased' group from the 'stayed the same' group at the four-digit industry level, but we can for the three-digit level (although only at the 10% level), it is ambiguous as to which relative measure might be considered 'better'. Further work is required here.

Perhaps a better assessment of the quality of the self-reported responses in Table 15 would come from testing whether any of the sub-group growth rates are significantly different from zero. From this perspective, two apparent facts are quite reassuring: first, none of the 'stayed the same' groups have mean growth rates significantly different from zero. Second, the 'increased' groups often have significant positive growth rates (for sales, productivity & four-digit market share). Conversely, the 'decreased' group is often indistinguishable from the 'stayed the same' group. Taken together these results suggest that combining low and middle-response groups and comparing them to the high group is an effective use of the qualitative data (as was done in Fabling & Grimes, forthcoming[19]).

5 Measurement Dissonance and Management Quality

In this section, we consider the implications of dissonance between subjective and objective measures of firm performance and management quality. We do this because it might be argued that one of the key abilities of people with a management role is to understand the environment in which they operate. If that is the case, one could use the distance between the subjective interpretation of the business environment and its objective measurement as a measure of management quality in this particular dimension. Two caveats should be borne in mind. First, there are many other reasons for there to be such a dissonance between perception and reality.

[18] Although the difference between the means of the 'decreased' and 'increased' is significant at the 1% level (F=8.57).

[19] Fabling & Grime's analysis was of Business Practices Survey (BPS) 2001 data. BOS 2005 is modelled on BPS with almost identical qualitative survey questions. The main difference between the surveys is that BPS does not contain as much quantitative data. The BPS is also part of the LFPD.

For example, the respondent may not have a role where they have access to the appropriate information (an accountant or financial manager may not be as aware of the business conditions in the market in which the firm operates as a manager with a more strategic role, and vice versa), or the respondent may have a different perception of what constitutes the appropriate market.

Second, understanding the business environment is only one dimension of the complex concept of 'management capability'. For example, should we call a manager who knows precisely how badly their firm is failing a good one for seeing this, or a bad one for presiding over it?

However, the area of measuring management capability is a difficult one and capability is often deduced from observed business practices. Ultimately, the usefulness of a measurement dissonance concept is in whether this information has any predictive power. It is beyond the scope of this paper to model the determinants of firm performance, but a brief exploratory analysis is perhaps appropriate at this juncture.

The first problem one encounters is the multiplicity of potential variables with which to generate a dissonance measure. We choose a single one: relative productivity. We attempt to use the (BOS-based) productivity of a firm to predict the relative productivity group into which the firm will place themselves. We do this using an ordered probit model with the dependent variable being a three category variable with 0 = 'lower than competitors', 1 = 'on a par with competitors' and 3 = 'higher than competitors'. The results of this exercise are presented in Table 16. The log labour productivity is found to be a statistically significant predictor of self-reported productivity group, although the model finds it difficult to differentiate between the 'lower than competitors' and 'on a par with competitors' groups (as shown by the standard error on the first cut point).

Table 16 Ordered Probit Regression of Relative Productivity

Number of obs	3,175	F(1, 3175)		9.3
Population size	34,760	Prob > F		0.0023
	Coef.	s.e.	t	P>t
ln(*LProd*)	0.1440	0.0472	3.05	0.002
cut1	-0.1880	0.4895	0.38	0.701
cut2	1.9046	0.4927	3.87	0

- *Observations rounded to nearest five for confidentiality reasons*
- *Weighted and stratified*

When we look at the predictions from the regression (Table 17), we see that no firm is actually predicted as having productivity that is 'lower than competitors'. A simple test of proportion of predictions would give a score of almost 60%. However, this is in large part due to the high proportion of firms in the 'on a par with competitors'. Indeed, if we had made the assumption that all firms were in this group, we would have done better! Our model has over-predicted firms in the 'higher than competitors' group.

Table 17 Predictions from regression

	Predictions versus actual values (% of total)			
Predicted group	Lower	On a par	Higher	Total
Lower than competitors	0	0	0	0
On a par with competitors	6.1	58.7	33.6	98.4
Higher than competitors	0.1	0.8	0.7	1.6
Total	6.1	59.5	34.4	100

- *Correct predictions = 59.5%, incorrect predictions = 40.5%*
- *Mean and Linearised Standard Errors based on weighted and stratified data*
- *Weighted and stratified*

The poor predictive power of our model may be the reason why we cannot differentiate between the firms who do or do not undertake 'good' business practices (Table 18). In the table, we group firms into those whom we correctly predicted their self-reported productivity group (labelled 'Right') and those for whom we did not ('Wrong'). For the two practices we consider here (how far ahead the business plans and whether there is someone in the firm who regularly assess whether the business is achieving its goals), we cannot differentiate between those for whom we predicted correctly and those who we did not. Our extremely simplistic model of management is rather a failure.

However, our model *is* very simplistic. We have not accounted for sample selection (i.e. the 'don't know' category), there are no control variables, and we have not modelled the complex interaction between management quality and firm performance. This latter point is of particular importance. There are many reasons to expect management quality and firm performance to be correlated. Clearly, if managers can affect firm performance, then we would expect them to be working in successful firms. Last year's successful firms are likely to also be this year's successful firms and part of the reason for that is the managers who made them

successful. Further, more successful firms can afford to bid higher wages in order to attract high quality managers.[20]

Table 18 Dissonance and Business Practices (%)

				Test of equality	
	Wrong	*Right*	*Total*	*F*	*p*
How far ahead business plans					
Up to 6 months	3.86	7.91	11.77	5.98	(0.4800)
Up to a year	13.49	20.95	34.44	14.205	(0.5970)
Up to 2 years	9.52	12.38	21.91	7.62	(0.5804)
More than 2 years	8.18	11.96	20.13	7.87	(0.1866)
Don't Know	0.44	1.19	1.63	0.97	(0.2233)
No goals set for this business	2.74	7.38	10.12	6.01	(0.2956)
Total	38.22	61.78	100	42.67	(0.0206)
Someone regularly assesses whether business is achieving goals					
Yes	22.39	36.31	58.69	25.115	(0.2367)
No	12.03	20.12	32.15	14.105	(0.6573)
Not applicable	3.8	5.35	9.15	3.45	(0.2558)
Total	38.22	61.78	100	42.67	(0.9663)

- *Weighted and stratified*

6 Conclusions

We have compared a number of subjective and objective measures of firm performance drawn from BOS (a survey with very high response rates) and IRD tax returns. There is much commonality in the picture we see using either administrative (tax) or quantitative survey data, giving us some comfort that the tax data, while not collected for statistical purposes, serves us well as a tool for measuring firm performance. This is not a trivial result – survey data may be considered superior to tax data because questions are designed to collect the right conceptual variable while, conversely, tax data may be considered superior because, for example, firms could be made subject to audits with penalties for inaccurate filing. The fact that we find good concordance across our quantitative data sources suggests either that these pros and cons balance out or that, more likely, the data comes from the same source (financial accounts). Demonstrating the usefulness of the tax data for economic research enables us to confidently

[20] A related point is the complementarity between skill and capital – firms with higher levels of capital tend to have higher labour productivity and skills are often complementary with capital.

construct longitudinal performance measures for our (cross-sectional) survey respondents (see Fabling et al. 2007 for more on this subject).

In our comparison of qualitative and quantitative performance metrics we were concerned with how sceptical we should be of subjective self-reported measures of firm performance. Concerns arise primarily because respondents do not have full information (eg, they cannot accurately estimate the performance of competitors), and because individuals may prefer to consider themselves in a more positive light than is perhaps warranted.

Overall, our assessment of the qualitative performance metric provides some comfort to those analyses that have relied purely on qualitative measures of firm performance. At least at the mean, our quantitative and qualitative measures of firm performance do appear to be internally consistent. However, it seems likely that the qualitative measures disguise a substantial portion of the signal, particularly where it comes to identifying performance differences in firms that sit outside the 'high' performance group. Common sense suggests that, given a choice, quantitative data should be preferred over self-reported figures, and our future work using this data will rely mainly on such metrics.

However, we believe there is still potential in the self-assessed performance metrics to yield insights into firm behaviour. Any "rose-tinting" tendency is not likely to be conducive to doing good business and so the possibility arises for any dissonance between perception and reality to be used as a measure of management capability. Our illustrative, back of the envelope, attempt at this shows that whilst this may be worthy of pursuit in the future, one must be careful how one frames such attempts.

7 Bibliography

Bertrand, Marianne, and Mullainathan, Sendhil, (2001), 'Do People Mean What They Say? Implications for Subjective Survey Data', *American Economic Review*, 91.2, pp. 67-72.

Cooper, A. C., C. Y. Woo, and W. C. Dunkelberg (1988), "Entrepreneurs' Perceived Chances for Success", *Journal of Business Venturing*, 3, pp. 97-108.

Cox, Val (2006), "IR10 purchases – further analysis", mimeo, Statistics New Zealand

DeBondt, Werner F.M., and Richard H. Thaler, (1995), 'Financial decision making in markets and firms: A behavioral perspective', in R. A. Jarrow, V. Maksimovic, and W. T. Ziemba, (ed.): *Handbooks in Operations Research and Management Science*, Volume 9, Finance . pp. 385–410 (Elsevier).

Fabling, Richard, (2007a), 'The feasibility of producing official statistics from the Longitudinal Performance Dataset', Statistics New Zealand, mimeo.

Fabling, Richard, (2007b), 'Just how innovative are New Zealand firms? Quantifying & relating organisational and marketing innovation to traditional science & technology indicators', Occasional Paper 07/01, Ministry of Economic Development, New Zealand.

Fabling, Richard, and Arthur Grimes (forthcoming), 'Practice makes profit: Business practices and firm success', Small Business Economics

Fabling, Richard, Arthur Grimes, Lynda Sanderson and Philip Stevens, (2006), 'Some Rise by Sin, and Some by Virtue Fall: Firm Dynamics, Market Structure and Performance', Presentation to New Zealand Association of Economists Annual Conference, Christchurch.

Forth, John, and McNabb, Robert, (2006a), 'Innovations in WERS 2004: The Collection of Objective Data on Workplace Performance', WERS 2004 Information and Advice Service Technical Paper no. 1.

Forth, John, and McNabb, Robert, (2006b), 'Workplace Performance: A Comparison of Subjective and Objective Measures in the 2004 Workplace Employment Relations Survey', mimeo, National Institute of Economic and Social Research.

Gervais, Simon, and Itay Goldstein, (2005), 'The Effects of Biased Self-Perception in Teams', Working paper, Duke University and The Wharton School.

Glaser, Markus & Weber, Martin, (2003), 'Overconfidence and Trading Volume', Sonderforschungsbereich 504 Publications 03-07, Sonderforschungsbereich 504, Universität Mannheim.

Hillage, J., Regan, J., Dickson, J. and McLoughlin, K. (2002), 'Employers Skills Survey 2002', Research Report 372, London: Department for Education and Skills.

Larwood, L., and Whittaker, W., (1977), 'Managerial Myopia: Self-Serving Biases in Organizational Planning', *Journal of Applied Psychology*, 62, pp. 194-8.

Mason, G., (2005), 'Enterprise product strategies, high value added production and employer demand for skills: methodological issues', mimeo, National Institute of Economic and Social Research, London.

Svenson, O., (1980), 'Are we all less risky and more skillful than our fellow drivers?', *Acta Psychologica*, 47, pp. 143-8.

Webb, Mike, (2007), 'Is overconfidence the X FACTOR in teamwork?', *Competition and Regulation Times*, March, p. 3.

8 Data Appendix

8.1 BOS Variables

We use SNZ-imputed values in case of item non-response. For quantitative variables used here, imputation is by stratum mean. For qualitative variables, imputation is nearest neighbour. In future versions of this paper we will drop imputed financial observations and treat qualitative non-responses as a separate "response" category.[21] As it stands, roughly 4% of weighted qualitative responses are imputed.

Objective Data

Employment

Employment is the sum of full-time working proprietors and employees plus one-half of the sum of part-time working proprietors and employees, where full-time is defined as working 30 hours or more per week. These data come from the answers to question 30 'At the end of the last financial year, how many staff worked for this business?' respondents are asked to include 'those temporarily absent from work (e.g. sick, on leave, strike or temporary lay off). Responses are recorded as data items A3001 (full-time working proprietors), A3002 (part-time working proprietors), A3011 (full-time employees) and A3012 (part-time employees).

Purchases

The figure for purchases comes from question 16: 'For the last financial year, what was the total amount this business paid for all other operating expenditure?' Firms are given the following examples: 'purchases of goods and services from suppliers' and 'renting and leasing costs'. They are asked to not include: salaries and wages; purchase of fixed assets; interest and finance costs; depreciation or amortisation; losses on sales of fixed assets. Data item A1601.

Sales

The figure for firm sales comes from question 8: 'For the last financial year, what was the total this business received from the sale of goods and services?' Data item A0801.

Value added

Value added is calculated as sales of goods and services (A0801) minus purchases (A1601).

[21] Imputation indicators for BOS are not currently available in the Datalab, but will be included soon.

Operating Revenue

Operating revenue is the sum of receipts from sale of goods and services (Question 8, data item A0801) and other operating revenue (Question 9, data item A0901). Other operating revenue includes renting and leasing income, government grants received for operating purposes and interest and dividend revenue. It excludes proceeds from the sale of fixed assets and the gains on the sale of fixed assets.

Operating Expenditure

Operating expenditure comes from the response to question 17. It is the sum of salaries/wages, interest payments, depreciation and amortisation, and all other operating expenditure. Other operating expenditure includes the purchase of goods and services from suppliers and renting and leasing costs. It excludes losses on sales of fixed assets. Data item A1701.

Operating Profit

Profit is defined as operating revenue less operating expenses.

Subjective Data

Relative Productivity and Profitability

Question 38. 'Mark one oval for each item listed. How do you think this business compares to its major competitors on each of the following?'

	lower than competitors	on a par with competitors	higher than competitors	don't know
profitability				
productivity				

Data items A3801 and A3802.

Performance change: sales, profitability, productivity & market share

Question 39. ' Mark one oval for each item listed. Over the last financial year, did the following items decrease, stay the same or increase for this business?'

	decrease	stay the same	increase	don't know
total sales of goods and services				
profitability				
productivity				
market share				

Data items A3901, A3902, A3903 and A3904.

8.2 Business Activity Indicator (BAI) Data

The Business Activity Indicator uses GST data from the Inland Revenue Department matched to the SMZ Business Frame. The BAI data come from the *Goods and services tax return form*, GST 101.

Sales

The sales data in the BAI relate to 'Total sales and income for the period (including GST and any zero-rated supplies).' This is adjusted using data on zero-rated sales as follows

$$S_E = \frac{8}{9}(S_I - Z) + Z$$

where S_E = Sales excluding GST, S_I = Sales including GST, Z = zero rated sales.

In a small number of cases zero-rated GST data is missing. This scenario arises when GST total sales (and purchases) have been imputed. For these observations, we assume Z equals zero in the GST adjustment process.

Purchases

The purchases data in the BAI also come from the *Goods and services tax return form*, GST 101. They relate to 'Total purchases and expenses (including GST) for which tax invoicing requirements have been met' as include an estimate for imported goods and the use of private goods and services in taxable activity adjusted by 8/9.

8.3 IR10 Data

Sales

The sales data recorded in the IR10 form relate to 'Gross income from sales and/or services' and are GST exclusive.

Profit

The profits data recorded in the IR10 form relate to total current year taxable profit.

8.4 PAYE Data

Employment

Employment is measured using an average of twelve monthly PAYE employee counts in the year. This is known as Rolling Mean Employment (RME). One problem with this data is that

it does not include working proprietors not drawing a wage subject to PAYE deduction. Because of this, following Fabling (2007b), we add one to the employment count of every firm in the dataset. Fabling *et al.* (2007), use the recent addition of a working proprietor count to the LFPD, and this approach will be followed in future work.