

INCENTIVES AND MANAGERIAL EXPERIENCE IN MULTI-TASK TEAMS: EVIDENCE FROM WITHIN A FIRM

Rachel Griffith[†] and Andrew Neely[‡]
April 2008

Abstract: This paper exploits a quasi-experimental setting to estimate the impact that a multi-dimensional group incentive scheme had on branch performance in a large distribution firm. The scheme, which is based on the Balanced Scorecard, was implemented in all branches in one division, but not in another. Branches from the second division are used as a control group. Our results suggest that the balanced scorecard had some impact, but that it varied with branch characteristics, and in particular, branches with more experienced managers were better able to respond to the new incentives.

Acknowledgements: The authors would like to thank the anonymous firm for use of the data and Laura Abramovsky, Rupert Harrison, Gareth Macartney, Imran Rasul, Helen Simpson and seminar participants at Cass Business School, London Business School and UCL for helpful comments. The analysis contained in this paper was funded by the Economic and Social Research Council (ESRC) and Engineering and Physical Sciences Research Council (EPSRC) through AIM. All errors and omissions remain the responsibility of the authors.

JEL classification: J33, M12, M52

Keywords: incentive design, balanced scorecard, managerial experience

Correspondence: rgriffith@ifs.org.uk; a.neely@cranfield.ac.uk

[†] Institute for Fiscal Studies and University College London.

[‡] Cranfield School of Management and AIM

1 Introduction

Many organisations provide some form of incentive pay to managers and workers. What form should these incentives take? This is the topic of many papers in the economic and management literatures. A well known theoretical result (Holmstrom and Milgrom, 1991) suggests that workers should not be made jointly responsible for single tasks, because sharing responsibility increases the total risk that each worker faces of successfully completing the task without increasing the benefit. Holmstrom and Milgrom also suggest that tasks should be grouped together, based on the cost of measuring and rewarding performance. Some workers should do the easy-to-measure tasks, and their pay should be contingent on performance, while other workers should focus on hard to measure tasks and receive fixed wages. This is because if a worker has both easy and hard to measure tasks they will concentrate on the easy to measure tasks, at the expense of the hard to measure tasks.

However, one of the most widely used measurement and incentive schemes - the Balanced Scorecard - runs counter to these results. The Balanced Scorecard was introduced by Kaplan and Norton (1992), and variants of it have been adopted by up to 70% of firms and organisations across the globe.¹ The idea behind the Balanced Scorecard is that managers can improve performance by monitoring and rewarding a range of linked activities (usually 15-20 covering four different perspectives – financial, customer, internal and innovation/learning). These activities are chosen to reflect both current actions that impact future performance, as well as the outcomes of past actions. Performance is typically evaluated at the group level (for example, business unit, branch or team) and incentive payments are often based on group performance. As well as forming the basis for an incentive scheme, the Balanced

¹ The most recent evidence is contained in Neely, A.D., Yaghi, B. and Youell, N. (2008). Other data are available in the Balanced Scorecard Hall of Fame at http://www.bsc.org/pdf/BSCHoF-Membersby_Industry_2000-2005.pdf, or see Gates (1999) who surveys 113 “leading” US, European and Asian companies and finds that 81% of respondents use a strategic performance measurement system; Maisel (2001) samples 1990 US management accountants and finds that 47% of respondents use a strategic performance measurement system; Rigby (2001) surveys 214 North American firms and finds that 44% of organisations use the Balanced Scorecard; Speckbacher et al (2003) estimate that 26% of firms in Germany, Switzerland and Austria use the Balanced Scorecard, and Marr et al (2004) find that 35% of North American organisations use it.

Scorecard also provides information to managers and workers: “*The Balanced Scorecard is like the dials in an airplane cockpit: it gives managers complex information at a glance.*” (Kaplan and Norton, 1991, p71)

As well as running counter to results from the economics literature, the use and effectiveness of the Balanced Scorecard has been criticised in the management literature.² In particular, Jensen (1991) argues that the Balanced Scorecard will do nothing to solve the agency problem. It provides no information on how managers and workers should trade off different objectives, so will fail as an incentive mechanism, although he accepts that it may provide useful information to managers, for example on the company’s strategy and drivers of value.

With all these arguments against the Balanced Scorecard why have so many firms adopted it? Has it been an effective incentive scheme? Does it provide useful information to managers? The rapid uptake of the Balanced Scorecard has been largely fuelled by high profile success stories in other firms, yet there is little rigorous empirical evidence on how well the scheme works.³

In this paper we exploit a quasi-experimental setting within a single firm in order to investigate the effectiveness of the Balanced Scorecard. The firm implemented the Balanced Scorecard in one division with the explicit objective to see how well it would work before rolling it out across the firm. This allows us to overcome one of the main problems in evaluating incentive schemes – the fact that who adopts the incentive mechanisms is endogenous - by comparing monthly performance across a large number of branches in two different divisions within the same firm.⁴

² See, inter alia, Cools and van Praag (2003), Fink (2004), Gosling (2003), Jensen (1991, 2001) and Ittner, Larcker and Meyer (1998), Nørreklit (2000), Nørreklit (2000, 2003).

There is also an earlier literature, for example, Ridgway (1956) criticises the idea of multidimensional incentives schemes.

³ Hoque and James (2000) survey 66 Australian manufacturing firms, Banker, Potter and Srinivasan (2000) look at 18 hotels, Malina and Selto (2001) consider multiple divisions of a large firm, Ittner, Larcker and Meyer (2003) conduct a case study in a large firm, Neely, Martinez and Kennerly (2004) consider 35 branches in one firm, Davis and Albright (2004) look at nine branches of a firm, Burgess et al (2004) evaluate a randomised trial incentive scheme in Job Centres in the UK.

⁴ Burgess et al (2004) are able to evaluate the introduction of a team based multi-task incentive scheme in Job Centres in the UK, where implementation was randomised.

Another advantage of our setting is that the firm initially used profit related pay in all divisions. This means that we are able to identify the impact of the particular form of incentive scheme from the general impact of any form of incentive pay.

We find that behaviour changed in response to the Balanced Scorecard, but overall this change did not lead to increased trading profits because while sales increased, costs also increased, and in some branches the increase in costs was greater. However, the impact varied across branches. We use interviews with a number of individual managers to gain a better understanding of why this heterogeneity in response arose. These interviews highlight the complexity of the Balanced Scorecard, and offer the explanation that experienced managers were better able to interpret the large number of indicators than less-experienced managers, and thus they were able to effectively allocate effort within the branch. We find empirical support for this.

The idea is that it is not only the incentives that matter, but also the ability of managers and workers to respond to them. When it is necessary for managers and workers to perform a large number of tasks (for example, the tasks involved in running a retail establishment), it is important that the manager can effectively decide where best to put both his and workers' marginal effort. The Balanced Scorecard gives the manager additional information on past performance, but does not tell the manager where additional marginal effort will be most effective. It gives the same incentive to all measures. This has been one of the main features that has been criticised (Jensen, 2001), that the Balanced Scorecard does not tell managers where marginal effort will be the most effective in improving performance. We interpret the fact that more experience managers were able to achieve better performance under the Balanced Scorecard as showing that it requires additional ability (acquired through experience) for the manager to know where increased effort will yield the greatest payoff.

The structure of the rest of the paper is as follows. In the next section we describe the firm, the Balanced Scorecard, how it was implemented and the impact we would expect to find. Section 3 describes the data and our empirical approach. Section 4 presents the results. A final section summarises and concludes.

2 The setting

We describe the firm, the incentive scheme, how it was designed and implemented, and then discuss what impact we expect to see on performance.

2.1 *The firm*

The firm is a multinational distributor of heating and plumbing products. It has thousands of branches in over ten European and North American countries and employs around 50,000 people. In the UK there are four main divisions. Both divisions have several sub-divisions, or brands. We use data on the largest brand in the two largest divisions. Each brand is made up of a large number of branches. During the period we consider branches dealt primarily with one brand, and acted as relatively small trading units, employing between 2 and 32 staff (with a mean of 10).

We compare performance in the dominant brand in Division 1 (where the Balanced Scorecard was introduced) with the dominant brand in Division 2. Prior to August 2002 both divisions used the same incentive scheme, which was based only on branch profits. Division 2 kept this scheme after August 2002. The two divisions are similar in terms of average sales and profits per branch. They differ in that they sell distinct products, but both sell products that are used in the construction industry. Does a Division 2 branch represent a good control group for a Division 1 branch? The goods that these firms sell are bought by the same customers - builders who buy products from Division 1 will typically also buy products from Division 2 - so the branches will experience similar demand shocks. In fact, since undertaking this analysis the firm has started selling both sets of products in the same store. The branches operate in similar economic conditions, for example, they employ similar people from the same labour markets.

The largest part of firm profit comes from individual branch profits, though orders are also received at head office and head office also receives some volume discounts from suppliers. In this study we focus on branch profits. Profits of an individual branch are the revenue earned on the sale of each product, minus the costs of sale, minus central branch costs. The main elements of cost at the branch level are the cost of goods sold,

labour costs, infrastructure (including capital costs), distribution and transport costs. Other costs include general and administration costs, spending on information technology, local marketing, advertising and other branch level administrative costs. These are described further in the data section below.

Effort of branch staff and branch managers can affect outcomes in the following ways:

- the price paid for a product can vary with each transaction (branch staff have discretion to negotiate individual prices, starting from a base price);
- staff deal directly with customers and thus have influence on the quantity, type and range of products sold;
- branch (as well as regional) managers influence the quantity sold through setting base price levels, marketing and setting special offers;
- staff and branch (as well as regional) managers' actions can affect hiring and firing costs (through staff retention levels), volume discounts (through the type and quantity of goods sold) and various other branch level costs in a number of ways.

A typical branch is managed by a branch manager,⁵ and has an administrator who works in the office, two or three sales staff, a driver, and three or four people who work in the warehouse or stockyard. Further details are given in Table A.1 in the Data Appendix. The role of the manager, among other things, is to decide on hiring decision (in conjunction with head office), allocate staff to tasks, decide on special offers and decide on the level and type of local marketing activity.

2.2 The incentive scheme

Prior to August 2002 employees in both divisions received a bonus that was a function of branch level profits. The bonus was allocated to branches based on a

⁵ The branch manager reports to a regional manager who reports to a brand operations director, who reports to a brand managing director who reports to a divisional managing director, who reports to the board.

percentage of their profits, and was allocated to branch staff by the branch manager, at his discretion and in consultation with the regional manager.

Senior management became concerned that the profit-based bonus scheme was leading to dysfunctional behaviours, as is emphasised in the literature.⁶ For example, managers had incentives to adjust the timing of capital investment where they would affect bonus payments, and they faced incentives to compete for business with other local branches, because profits were calculated on a point of sale basis.⁷ The firm was also concerned that branch staff were not putting sufficient effort into activities that enhanced long-run profitability, such as maintaining customer loyalty and relationships with suppliers. Senior management decided to change the basis of the performance measurement and incentive scheme in the organisation. They decided to trial a new scheme, based on the Balanced Scorecard, in Division 1 prior to rolling it out across the firm. The Balanced Scorecard was designed to overcome the dual problems of subjective performance evaluation (which can give rise to various forms of bias and encourage workers to waste effort to curry favour with managers) and focussing on one key performance target (which can lead to dysfunctional behaviour with workers focusing all their energy on tasks that are rewarded, and ignoring those that are not included in the incentive scheme).⁸

The new incentives scheme made three big changes (i) it used multiple targets rather than a single target, (ii) it was non-discretionary rather than discretionary, (iii) it provided more information to managers on a wide range of indicators, such as customer behaviour and efficiency, than was previously available.

The specific incentive scheme that the firm implemented works as follows. The firm identified 17 key indicators of performance, including both financial and non-financial indicators of performance. Incentives were paid to each employee on a six-monthly basis, but effectively earned on a monthly basis. For individual branch staff and branch managers the payment was based on the total number of points the branch

⁶ See, *inter alia*, review by Prendergast (1999) and Hayes and Abernathy (1980).

⁷ See, *inter alia*, Oyer (2004) and Larkin (2006).

⁸ See Argyris (1952), Ridgway (1956), Holmstrom and Milgrom (1991), Kerr (1995) and for a recent review of the literature Prendergast (1999).

earned in the month (divided by the number of staff) times the value of a point.⁹ Each branch is graded “green”, “amber” or “red” on each of the 17 measures contained in the balanced scorecard. A green grade earns 3 points, an amber grade 1.5 points and a red grade 0 points. The total number of points earned by the branch is given by,

$$\text{points} = (3 \times \text{green} + 1.5 \times \text{amber}) \times \text{number of people in branch.}$$

There are 17 measures in total so the maximum number of points per employee a branch can earn in a month is 51. In 2003 a point was worth £1 for branch staff (the value of a point is higher for managers) so the maximum bonus a branch worker could earn in a month was £51, or £612 a year. Salaries for branch workers average around £12,000, so the maximum bonus represents around 5.1% of salary.

The total amount the firm allocated for the scheme was just under £2m. This was set aside in a separate account to signal the firm’s commitment to the scheme. The previous scheme, based on profits, cost about the same total amount.¹⁰ As well as changing the incentive structure, an important aspect of the Balanced Scorecard was the provision of more detailed information on performance, and crucially on several *leading* indicators of profitability. Each branch manager received a detailed report every month on the performance of their branch on each of the 17 measures.

Why did the firm implement the Balanced Scorecard in Division 1 and not Division 2? This is an important question, as our strategy for identifying the impact of the Balanced Scorecard relies on performance being independent of this decision. The pragmatic reason is simply that the Divisions were run relatively independently and the Managing Director of Division 1 was rather more keen on management innovation, and particularly the Balanced Scorecard. At the outset the organisation considered whether it would subsequently implement the Balanced Scorecard in other divisions, but in essence decided to delay a decision and instead treat Division 1’s implementation as a pilot study. Thus, we argue that the introduction can be treated as

⁹ For regional managers it is based on the average of the points earned in branches under their control. For central office staff it is calculated based on a simplified version of the balanced scorecard containing only measures that the central office staff could affect.

¹⁰ The firm estimates the cost of implementation at around £0.5m, including the direct and indirect costs such as management time.

independent of any expectations about the impact the Balanced Scorecard would have in one Division over the other.

2.3 The design and implementation process

In evaluating the impact of the Balanced Scorecard we need to be careful to distinguish two questions - (i) did the new incentive scheme change behaviour, and (ii) did this behaviour lead to improved performance. The idea behind the Balanced Scorecard is that the firm needs to determine what are the key drivers of future profitability and build these into the scheme.

The board devoted significant time to discussing what were the objectives for the division, what behaviours they wanted to encourage in the branch network, and how these behaviours might be reinforced through the choice and design of appropriate performance measures. At various stages during the process the directors consulted regional and branch managers. The firm invested considerably in education and training, as well as engagement with workers at all levels throughout this time.¹¹ In the end the firm adopted 17 measures, shown in Figure 1 and described further in Tables 1.¹²

[Figure 1 around here]

[Table 1 around here]

Following a 9 month design and deployment process the Balanced Scorecard was introduced in August 2002. The scheme remained in operation until July 2004.

2.4 Anticipated impact on performance

How do we expect the three key changes - (i) introduction of multiple measures, (ii) move from discretionary to non-discretionary, and (iii) increased information to affect

¹¹ The discussions were about what the objectives of the division were and who had control over them. Meetings were held in every branch and discussion was in depth and focussed on what individual workers could do to improve performance.

¹² Personal development was included as a measure, but the data were not adequately collected, so it was not used. The employee satisfaction survey had a low response rate (e.g 32% in December 2002), but the firm used it anyway.

performance? Remember that, as mentioned above, there was no major change in the overall value of incentives, just in their form.

In order to consider how it would affect performance we consider the incentives that individuals faced and how they changed. We assume that individuals seek to maximise their income, net of effort. Income consists of a base wage and an incentive payment, which is a function of performance. Performance is a function of the effort of all staff, managers and directors. Effort is costly. Individuals will exert effort up to the point where the marginal cost of effort equals the marginal benefit, in terms of the incentive payment. This is true under either scheme. What changed is the way performance is measured and rewarded.

The main objectives of the firm were to encourage workers to put more effort into a broad range of activities that were previously not rewarded, but which the directors of the firm believed feed into long-term profits (long term being over several months to one to two years). The tasks concerned are substitutes from the individual workers point of view (they each take time and an individual worker can spend time on one task or another), but from the point of view of the firm and value-maximisation they are complements, in the sense that workers need to spend time on all of the tasks in order to maximise the value of the firm. For example, a worker can either sweep the floor or restock the shelves - from the workers point of view these are substitute activities. But from a customer's point of view these are complementary characteristics of a shop - they want both a clean shop and well stocked shelves.

Where workers undertake tasks that are substitutes for each other, and where the rewards are equal and independent (as is the case here), then workers will devote more time to those tasks they find easiest. If the measurement of some tasks is more precise, or more clearly understood – in the sense that it is easier to identify improvements in performance - then workers will devote more time to those tasks that are measured more precisely.¹³ This is essentially the objection that Jensen (2001) and others have made to the Balanced Scorecard - it does not give clear guidance to workers as to what is the most important task to perform (should they spend another 10 minutes sweeping the floor, or should they start stocking the shelves), but rather

allows them to put too much effort into non-profitable tasks (either because they have mis-information about the payoff to the individual tasks or because both are rewarded so they simply do the one that is easiest).

The Balance Scorecard scheme with 17 measures was complex, and managers were provided with large amounts of information. In order to effectively use this information the manager needed to be able to assimilate it and understand what it meant in terms of future profitability. A key issue is that the data as presented in the Balanced Scorecard were very aggregated, hence managers were not able to directly take action based on them. Each manager needed to be able to translate the information provided on the Balanced Scorecard so that it told them something specific about the local situation. For example, consider one of the measures - customer retention. Each month the manager would be given information on how many customers they were retaining compared to the same time last year. This is a useful leading indicator of profitability. However, simply knowing that you retained 5% fewer customers than this time last year is not sufficient information to enable the manager or shop workers to act. They needed to be able to figure out which customers were not coming back and why in order to know where to put effort that would increase retention rates. This is where experience comes in. Our thesis, developed through interviews with branch managers, is that experienced managers are either more likely to have seen in the past, or are better able now to develop, local solutions that allow them to take action to improve performance according to the Balanced Scorecard. A more experienced manager will both be better able to interpret the large number of indicators, and better able to (e.g. more credibly) motivate staff to put effort into the activities that matter for performance; with an inexperienced managers workers can easily become overloaded with information and tasks, leading to underperformance. We explore this idea in our empirical analysis below.

¹³ See Holmstrom and Milgrom (1991) and Burgess et al (2004).

3 Data and econometric method

3.1 Data

The main data is drawn from the monthly Profit and Loss (P&L) accounts of the firm. We have information at the branch level on sales, gross profits (sales minus cost of goods sold), trading profits, labour costs, infrastructure, transport costs, general and administrative costs, information technology investment, local marketing and advertising expenditure and other costs. We have these data monthly from August 1999 to July 2005. We also have data on each employee in each branch in Division 1 including their age, job title, and length of tenure in the firm from August 2003 to July 2004.

We use the location (postcode) of each branch to match branches from the two divisions. There are a number of factors that affect sales, but that are both exogenous (not affected by actions of the firm) and will have a common effect across both divisions. These include the economic cycle, local economic and labour market conditions and other local factors. There may, however, be some variation in demand conditions. The products sold by Division 1 are used both inside and outside buildings, while those sold by Division 2 are predominantly used inside. Thus weather conditions will have a differential affect on demand for the two goods. To control for this we use monthly data from twenty-six weather stations in the UK on the minimum temperature (in Celsius) and rain fall (in millimetres). In addition, goods sold by division 2 are used more intensively in refitting houses, while goods sold in division 1 are used more intensively in new buildings. To control for this we use data on total quarterly construction activity in a range of categories to capture variation in aggregate demand. These data come from the Construction Products Association. We aggregate separate information on aggregate quarterly output for new private sector housing, private and public sector remodelling and commercial and industrial development. The firm has given us the weight of each of these in demand for each of their brands, and we use this to construct a measure of quarterly aggregate demand for each brand.

Tables 2 and 3 show descriptive statistics of the main variables. Table 4 shows descriptive statistics on the number of employees, types of jobs within a branch, and the average experience and age of staff.

[Tables 2, 3, 4 around here]

We match each branch from Division 1 brand A to the geographically nearest branch from Division 2 brand C. The markets for the goods sold by the firm are very local - the firm tells us that consumers are rarely willing to travel further than 30 kilometres. There are four Division 1 branches which do not have a Division 2 branch within 30 km, and we exclude these from our analysis. The average distance between matched branches is 4.5 kilometres.

3.2 *Econometric Method*

Our setting allows us to adopt a quasi-experimental design method¹⁴ and use a combined matching and difference-in-difference estimator. As highlighted above, a major problem in the literature that attempts to identify the impact of incentive schemes has been the fact that organisations choose whether and which incentive schemes to adopt - the adoption of the incentives scheme is endogenous.

We are able to use the fact that the firm implemented the Balanced Scorecard in one part of the firm, but not the other, to tackle this problem. Differencing between the matched branches allows us to control for all of those factors that have a similar affect on the two branches. However, the firm has highlighted two specific factors that may affect demand for the products sold in the two divisions differently, and we control for these in our analysis. We believe that this intra-firm inter-division comparison gives us a substantially better control group than is generally used in literature (where for example a different firm is generally used, where the differences between treatment and control will be much greater).

¹⁴ See, inter alia, Cook and Campbell (1979) and Heckman, Ichimura and Todd (1997). For a recent survey see Blundell and Costa-Dias (2005).

To formalise this, we can write the determinants of sales for each type of branch, where we denote branches within Division 1 with subscript i and within Division 2 with subscript j

$$(1) \quad \begin{aligned} S_{irt} &= \alpha_i + \beta_1 L_{rt} + \beta_2 D_{1t} + \beta_3 \tau_t + \beta_4 W_{rt} + \gamma BSC_t + e_{irt} \\ S_{jrt} &= \alpha_j + \lambda_1 L_{rt} + \lambda_2 D_{2t} + \lambda_3 \tau_t + \lambda_4 W_{rt} + \varepsilon_{jrt} \end{aligned}$$

where S : sales, L : local economic factors that affect costs such as labour markets, supply chain, or that affect local levels of demand, D : factors affecting national demand, τ : seasonal (monthly) dummies, W : weather, BSC : indicators of the use of the balanced scorecard, e, ε : idiosyncratic shocks.

We match each branch from Division 1 to the geographically nearest branch from Division 2 and consider the difference between the two branches to give us,

$$(2) \quad \begin{aligned} (S_{irt} - S_{jrt}) &= (\alpha_i - \alpha_j) + (\beta_1 - \lambda_1)L_{rt} + (\beta_2 D_{1t} - \lambda_2 D_{2t}) \\ &\quad + (\beta_3 - \lambda_3)\tau_t + (\beta_4 - \lambda_4)W_{rt} + \gamma BSC_t + (e_{irt} - \varepsilon_{jrt}) \end{aligned}$$

We have argued that $\beta_1 = \lambda_1$, i.e. local market conditions effect the two type of branches in the same way, so that they drop out of the difference equation, and that $\beta_2 = \lambda_2$, i.e. that changes in aggregate demand feed through into sales and profits in the same way, so that we can include the difference in aggregate demand for goods sold in the two branches. This gives:

$$(3) \quad \begin{aligned} (S_{irt} - S_{jrt}) &= (\alpha_i - \alpha_j) + \beta_2 (D_{1t} - D_{2t}) + (\beta_3 - \lambda_3)\tau_t \\ &\quad + (\beta_4 - \lambda_4)W_{rt} + \gamma BSC_t + (e_{irt} - \varepsilon_{jrt}) \end{aligned}$$

which is a form of combined matched and difference-in-differences estimator.

To investigate the idea that there was a heterogeneous impact of the introduction of the Balanced Scorecard we extend this specification to allow the impact of the Balanced Scorecard to vary with other observable characteristics of the branches, in particular the level of experience of senior staff, $\gamma = \gamma_0 + \gamma_1 Experience_i$.

4 Results

4.1 *The overall impact of the balanced scorecard*

We start in Table 5 by considering how we parameterise the BSC variables. The dependent variable is the difference in sales between the matched branches. In the first column we include an indicator for just the two years that the incentive scheme was in place (August 2002 to July 2004) - so we are comparing the level of sales in Division 1 branches with Division 2 branches during this period with the difference in the level before and after this period. This shows that sales in Division 1 branches increased by an average of £4,538 per month more than Division 2 branches during the two year period in which the Balanced Scorecard was in place. In column (2) we split this period in half and see that the impact was similar across the two years, although somewhat larger in the second year. In column (3) we also include an indicator for the year after the firm put the incentive scheme on hold - so we are now comparing only to the two years before the Balanced Scorecard was implemented. We now see a larger impact from the Balanced Scorecard, with a decline after the scheme was put on hold, though perhaps not immediately back to the original levels. In column (4) we consider whether there was an anticipation effect in the initial discussion period,¹⁵ but find no evidence of this. For the remainder of this paper we stick to the parameterisation in column (3). In Table 5 we also see that demand for Division 2 products grew somewhat faster than Division 1 products (the coefficient on demand is negative), that there is more demand for Division 1 products when it is hotter, and less when it is raining.

[Table 5 around here]

These results suggest that the Balanced Scorecard had an impact through increasing sales, but what about profits? In Table 6 we look at how gross profits, trading profits, labour costs, infrastructure expenditure, transport costs, general and administrative costs, information technology expenditure, local marketing and advertising expenditure and other costs were affected.

[Table 6 around here]

The first column of Table 6 repeats column (3) of Table 5. All regressions include controls for aggregated demand, the weather, month and branch effects. In column (2) we see that there was also a corresponding increase in gross profits (the difference between sales and gross profits is the cost of goods sold). In column (3) we see that this did not feed through into an increase in trading profits, and in fact led to a fall in profits in the later period of on average -£1,018. The difference between gross and trading profits is the costs considered in columns (5)-(10). The fall in profits arose largely because transport costs (column 6) and labour costs increased (column 4), and to a lesser extent because infrastructure (column 5) and general and administrative costs (column 7) increased. Other and IT costs actually fell.

This is the first empirical result of the paper - the Balanced Scorecard lead to an increase in sales, but costs increased by at least as much as sales at the aggregated branch level, so there was no increase (and actually some decrease) in profits. We emphasise these results in Figures 2 and 3.

[Figure 2 around here]

To produce Figure 2 we run a regression similar to that shown in column (1) of Table 6 but with a separate time dummy for each month.¹⁶ Figure 2 shows a plot of the monthly dummies smoothed over the period shown. The figure shows that prior to the introduction of BSC in August 2002 (200208) the difference in sales between Division 1 and Division 2 Branches was fairly constant. After the introduction of the BSC sales in Division 1 branches grew more rapidly. When the BSC was put on hold the difference in the level of sales evened out again.

[Figure 3 around here]

To produce Figure 3 we run a similar regression but now with the difference in trading profits as the dependent variable. The figure plots the monthly dummies smoothed over the period shown. The figure shows that prior to the introduction of

¹⁵ What is often called a Hawthorne effect in anticipation of the actual implementation.

¹⁶ I.e. a regression of the difference in the level of sales between each Division 1 branch and its matched Division 2 branch regressed on branch fixed effects, seasonal fixed effects, weather and time dummies.

BSC in August 2002 the difference in trading profits between Division 1 and Division 2 Branches was declining. After the introduction of the BSC trading profits in Division 1 stopped falling, but they did not grow in the same way that sales did (this is because a number of large components of costs grew more rapidly in Division 1 branches than in Division 2 branches). When the BSC was put on hold the difference in the level of trading profits did not change.

4.2 Did the balanced scorecard target the right tasks?

One important question is whether the Balanced Scorecard targeted the right tasks - i.e. tasks that would improve performance. Did branches that improved on the non-financial aspects measured, subsequently experienced improved performance? Did the Balanced Scorecard correctly identify the key drivers of performance? We use the data collected under the Balanced Scorecard to distinguish between branches that successfully put effort into improving on the non-financial measures from those which did not. For each branch we calculate what proportion of the total non-financial points that were available they earned. This ranges from 16% to 74% and has a median value of 52%.

In Table 7 we split the sample into those branches that were below and above the median share of points earned on the non-financial measures. In the left-hand side of Table 7 we consider the change in outcomes of those branches that either did not try, or were not able, to perform well on the non-financial measures, and on the right-hand side those branches that did well on the non-financial measures. We focus on sales and trading profits.

[Table 7 around here]

We see large differences between the two groups. In the left-hand panel branches that did poorly on non-financial measures did not experience any significant growth in sales, but they did experience some increases in costs, leading to *reduced* trading profits. In contrast, those branches that did well on the non-financial measures saw an increase in sales and trading profits. While costs did increase in these branches, sales increased by more, resulting in higher profits. Figure 4 shows a similar picture. Each dot represents a branch. The x-axis shows the share of possible non-financial points

earned, and the y-axis shows branch profits (scaled by sales). The correlation between these is 0.66 and is statistically significant.

[Figure 4 around here]

These results provide some evidence that the Balanced Scorecard measures were well chosen - those branches that did well on the non-financial performance indicators also did well financially. But, one concern is that this could simply be showing that good branches (or good branch managers) do well on both financial and non-financial indicators. Proponents of the Balanced Scorecard argue that giving branches an incentive to put effort into a broader range of factors that feed into long run performance will lead to better performance in the long run. Opponents of the Balanced Scorecard argue that giving managers and workers so many different incentive will lead to worse performance as individuals lose focus and put too much effort into the easiest tasks. To investigate this we make a further comparison. We consider branch financial performance prior to the Balance Scorecard (August 1999 - July 2002) and during the Balanced Scorecard (August 2002 - July 2004). We split the sample by financial performance in the period before the Balanced Scorecard was introduced.

[Figure 5 around here]

In Figure 5 we show that good performance on the non-financial points during the Balanced Scorecard led to good financial performance afterwards in both cases, and the relationship is stronger in those branches which previously performed badly than in those that did well (correlation coefficient (p-value) of 0.727 (0.000) and 0.346 (0.002) respectively).

These pictures are suggestive, but we haven't controlled for any of the local economic time varying factors we were concerned with above. To do this we return to the regressions and we compare performance in four groups of branches - (1) those who did badly on financial measures prior to implementation of the Balanced Scorecard, and who subsequently also did badly on non-financial measures during the implementation of the Balanced Scorecard, (2) those who did badly on financial measures prior to implementation and well on non-financial during implementation, (3) those who did well on financial measures prior but badly on non-financial during,

and (4) those who did well on financial prior and well on non-financial during. We focus on the results for sales and trading profits.

The far left-hand panel of Table 8 shows the estimates for group (1), the second panel for group (2), the third panel for group (3) and the far right-hand panel for group (4).

[Tables 8 around here]

What we see is that branches that do well on the non-financial measures also do well on financial measures, regardless of whether their financial performance was above or below average before the introduction of the Balanced Scorecard. In particular, those branches that previously did badly on financial measures (second panel from the left in Table 8) had substantially higher sales after implementation of the BSC, and some increase in trading profits.

Consider the branches that did well financially prior to implementation. Those that did badly on non-financial measures (third panel from the left in Table 8) did badly in financial terms after implementation, while those that did well on non-financial measures (far right panel) did well in financial terms after implementation.

To summarise, these results suggest that those branches that were successful on the non-financial measures also did well financially. The results also suggest that there was heterogeneity in the effect of the Balanced Scorecard across branches. We now turn to investigate what might explain this heterogeneity.

4.3 *The importance of experience*

From a theoretical perspective why should the impact of the balanced scorecard vary by the experience of managers? Some studies have argued that less experienced managers rely on formal measurement systems to supplement their inexperience, more than experienced managers (Bourne et al, 2005). Empirically, we find the opposite, that the more experienced the manager the greater the impact of the balanced scorecard. As an incentive scheme there is no obvious reason why more experienced managers should be more motivated by the incentive scheme than less experienced managers (if anything, we might think the opposite as the incentive will account for a lower share of their total income). The consumer psychology literature

makes an important distinction between motivation, ability and opportunity (Batra and Ray, 1986; Andrews, 1988; MacInnis et al. 1991), which is also discussed in the organisation design and psychology literatures (Parker and Wall, 1998). This literature makes explicit the distinction between: *Opportunity* (the freedom and scope to act), *Motivation* (the incentives to respond), and *Ability* (whether an individual has the necessary skills and capabilities to respond). The point made in both the consumer psychology and organisation design literatures is that the absence of any of the above factors significantly limits the likelihood either of action or the impact of action. In the case of the balanced scorecard it is clear that there is an incentive for managers to act – hence the motivation is in place – but do they have the opportunity and the ability to act? And are the answers to these questions affected by their experience?

To investigate this we undertook 20 systematic telephone interviews with branch managers. The branch managers selected for interview had all been with the firm for at least four years, hence they had been in post for the entire period during which the Balanced Scorecard operated. Each interview lasted between 45-75 minutes, during which time branch managers were asked a series of questions about their experiences with the Balanced Scorecard. Initially the interviewees were asked to explain when they first heard about the Balanced Scorecard and what their initial reactions were to the scheme. Next they were asked to explain how they introduced the Balanced Scorecard to their staff and to comment on the staff's reaction to the scheme. Third the interviewees were asked to explain how they used the Balanced Scorecard and how they involved their staff in discussing the results. Fourth they were asked to comment on what happened to the Balanced Scorecard over the time it was in operation. Finally they were asked to comment on how they measured and managed performance in their branches today (post the Balanced Scorecard).

From the interviews we found that, while the majority of managers interviewed were positive about the scheme, several raised issues that lead us to question whether they were able to act on the data. The first common theme to arise was the issue of data quality: *“I think one of the problems straight away was we didn't have the ability to measure a lot of the measures correctly or the figures were wrong” – Branch Manager A.*

The second was whether the measures were under the manager's control.

“Personally, we would discuss why it had happened and we would put an action plan in, just a vehicle for action plan, what we could try and do to improve the score, but there were certain measures that you couldn't control like your cash collection. If a customer had no money, there was absolutely nothing you could do to encourage them to pay the bill, so that was totally out of your control” – Branch Manager B.

“Quite honestly, I used to look at the control card and I used to see reds, I used to see the yellows, I used to see the greens and I knew which ones I could influence easily. Okay, and I could do something about those and I would do it. The ones I couldn't do, I just had to leave” – Branch Manager C.

Taken together – these two issues – the quality of the data and the question of whether the measures were under the managers' control, clearly have an impact on a manager's opportunity and ability to act, as well as their motivation to act. Poor quality data undermines the managers' confidence in the scheme, causing them to question whether the scheme is having a positive effect.

Interestingly, the issue of control over the measures is a factor that many managers raised, but then some found solutions to. For example, one experienced manager, explained the local work around that he and his team had developed to enable him to act on the customer retention data that other managers said they had no control over.

“I mean the staff that I've got on the counter have been with me for years and they will come in and say we haven't seen Brian. If they tell you this bloke hasn't been in for a week or so, you can do something about it” – Branch Manager D.

The interview data suggest that the more experienced managers found ways of overcoming the constraints of the balanced scorecard. They did not allow constraints, such as the lack of detailed and actionable data, to hamper their ability to act. In essence their experience enabled them to identify ways of overcoming these constraints.

We wanted to explore the idea that more experienced managers were better able to utilise the information in the Balanced Scorecard. The number of interviews was not sufficient to allow us to analyse the response systematically. Instead, we used information from the firm's payroll system. We use information on the years of experience of staff in each branch to look at whether more experienced workers, and in particular more experienced managers, responded better to the Balanced Scorecard.

In Table 9 we allow the impact of the Balanced Scorecard to vary with the average years of experience of all staff and of senior staff. Before considering these results, we note that a simple correlation in the data is consistent with the idea that more experienced managers achieve better performance. The correlation between years of managerial experience and profits is 0.36 (significant at the 1% level).

Turning to the results, in the first and second column of Table 9 we see that experience matters. A branch with staff with the average years of experience (6.6 years) will have a higher level of sales after the Balanced Scorecard was introduced than a branch with all new staff. In the third and fourth columns we look at trading profits and show that it is the years of experience of senior staff that matters most. A branch with senior staff with the average years of experience (11.3 years) will have a higher level of trading profits after the Balanced Scorecard was introduced than a branch with all new senior staff.

In results not shown (available from authors on request) we include the average age of all staff and of senior staff interacted with the Balanced Scorecard indicators and show that it is experience, not age, that is important. Also in results not shown (available from authors on request) we show that if we simply split the sample on whether the manager has more or less than 10 years experience (approximately the median) we see all of the effect of the Balanced Scorecard in the greater than 10 years experience group.¹⁷

Because this is a key result of the paper we illustrate it further with two figures.

¹⁷ The coefficients (standard errors) on the Balanced Scorecard variables are for the branches with managers with *less* than 10 years experience: (Aug 2002-Jul2003) 206 (968); (Aug 2003 - Jul 2004) 70 (847); Aug 2004 - Jul 2005) -1850 (796); for the branches with managers with *more* than 10 years experience: (Aug 2002-Jul2003) 1854 (896); (Aug 2003 - Jul 2004) 2218 (846); Aug 2004 - Jul 2005) -1339 (709).

[Figures 6 and 7 around here]

In Figure 6 we plot the monthly dummies (smoothed over 12-months) from a regression of difference in the level of sales between each Division 1 branch and its matched Division 2 branch, including branch fixed effects, seasonal fixed effects and weather. The dotted line is the same as appears in Figure 1. The top line (long dashes) is for branches where the senior staff have above average level of experience. The middle dotted line show the smoothed dummies when we use all branches. The lower solid dark line shows the smoothed dummies when we estimate using just branches where the senior staff have below average level of experience.

The figure shows that prior to the introduction of BSC in August 2002 (200208) the difference in sales between Division 1 and Division 2 Branches was fairly constant for all groups, with branches with more experience senior staff having a higher relative level of sales. After the introduction of the BSC sales in Division 1 branches with more experienced senior staff grew more rapidly, while sales in branches with less experienced senior staff did not. When the BSC was put on hold the difference in the level of sales evened out in all groups.

Figure 7 repeats this exercise for trading profits. The figure shows that prior to the introduction of BSC in August 2002 (200208) the decline in the relative level of trading profits between Division 1 and Division 2 branches was in branches with less experienced senior staff. After the introduction of the BSC trading profits in branches with more experienced senior staff grew more rapidly, while trading profits in branches with less experienced senior staff continued to fall. When the BSC was put on hold the difference in the level of trading profits evened out in all groups.

5 Summary and conclusions

Our results suggest that the balanced scorecard had some impact, and that this impact varied significantly by branch. Sales increased on average across all branches, but costs increased by at least as much, so that while gross profits did increase, trading profits for the business as a whole did not increase. There is significant variation in the impact. When we separate those branches that perform well on non-financial measures from those that perform poorly on non-financial performance measures we

find that the first group – those that perform well on non-financial measures – experience statistically significant increases in sales, gross profits and trading profits. One potential explanation of this finding is that branches that perform well on non-financial measures and financial measures are simply well managed. Note that we are allowing for each branch to have a different average level of performance (we include branch fixed effects). To consider this further we split the sample into branches that perform relatively well and those that perform relatively poorly on financial measures in the pre-implementation period. We then explore whether there are differences in non-financial and financial performance in the implementation period. We find that, regardless of prior performance, branches that perform well on non-financial measures also perform well on financial measures. This finding is particularly important as it suggests that the balanced scorecard, when implemented correctly and adopted by the branches, has a positive impact on branch performance in terms of sales, gross profit and net profit.

We then show that years of experience is an important factor in explaining these differences in performance. More experience managers were able to improve performance. We interpret this as suggesting that the information content of the Balanced Scorecard is what is important, not the incentive per se.

Crucially the research reported in this paper suggests that multi-dimensional performance measurement systems can have a positive impact on a firm's financial performance providing the measures and the associated data are presented in a way that enables managers and staff to act on them. It is not enough to introduce an incentive scheme that relates to the measures, unless that incentive scheme is accompanied by data that are meaningful to those who have to manage and improve performance. Too often, when performance measurement systems and associated incentive schemes are introduced they are not designed to enable staff at the front line to take action. The schemes are too complex and contain data which are not sufficiently disaggregated to enable action to be taken. How to designing measurement and incentive schemes configured for action is a significant challenge for future research.

References

- Andrews, J.C. (1998) "Motivation, Ability and Opportunity to Process Information: Conceptual and Experimental Manipulation Issues," in Michael J. Houston (ed.), *Advances in Consumer Research*, Vol. 15 (Provo, UH: Association for Consumer Research), pp. 219–225.
- Baker, G (1992) "Incentive Contracts and Performance Measurement" *Journal of Political Economy*, 100, 598-614
- Baker, George P., Michael C. Jensen and Kevin J. Murphy, 'Compensation and Incentives: Practice vs. Theory,' *Journal of Finance* (July, 1988)
- Banker, R. Potter, G and Srinivsan, D. (2000) "An Empirical Investigation of an Incentive Plan That Includes Nonfinancial Performance Measures", *Accounting Review*, 75(1), 65-92.
- Batra, R. and Ray, M.L. (1986) "Situational Effects of Advertising Repetition: The Moderating Influence of Motivation, Ability and Opportunity to Respond," *Journal of Consumer Research* 12, pp. 432–445.
- Blundell, R and Costa-Dias, M. (2000) "Evaluation Methods for Non-Experimental Data" *Fiscal Studies* vol. 21, no. 4, pp. 427–468
- Blundell and Costa-Dias (2005) "Alternative Approaches to Evaluation in Empirical Microeconomics" UCL mimeo, <http://www.econ.ucl.ac.uk/courses/mect2/rbmcd-13-03-2006.pdf>
- Bourne, M.; Kennerley, M.; Franco, M. (2005) "Managing through Measures: A Study of the Impact on Performance", *Journal of Manufacturing Technology Management*, 16, 4, pp. 373-395.
- Burgess, S., Propper, C., Ratto, M. and Tominey, E. (2004) "Incentives in the Public Sector: Evidence from a Government Agency" *CMPO Working Paper* No. 04/103
- CMPO, <http://www.bris.ac.uk/cmpto/research/incentives/inctablehk.htm>
- Cook, T.D. and Campbell, D.T. (1979) "Quasi-Experimentation: Design and Analysis Issues for Field Settings", Houghton Mifflin.
- Cools, Kees and Mirjam van Praag (2003), "The Value Relevance of a Single-Valued Corporate Target: An Explorative Empirical Analysis." *Tinbergen Discussion Paper* TI2003-049/3, <http://www.tinbergen.nl/discussionpapers/03049.pdf>
- Davis, S. and Albright, T. (2004), "An investigation of the effect of Balanced Scorecard implementation on financial performance", *Management Accounting Research*, 15, 135-153
- Fink, R "New Carrots, Old Yardsticks?" *CFO Magazine* June 2004
- Gosling, J. (2003), *A Worldly Mindset: Satisfying the needs of stakeholders. Creating an Agenda for European Leadership - Conference Proceedings*, Bled 10-11 June 2003 <http://www.leadership-studies.com/research/viewabs.asp?id=42>

- Hayes, R.H. and Abernathy, W.J. “Managing Our Way to Economic Decline”, *Harvard Business Review*, July-August, 1980, pp. 67-77.
- Heckman, J, Ichimura, H. and Todd, P. (1997), ‘Matching as an econometric evaluation estimator’, *Review of Economic Studies*, vol. 64, pp. 605–54.
- Hoque, Z. and James, W. (2000), “Linking Balanced Scorecard Measures to Size and Market Factors: Impact on Organisational Performance”, *Management Accounting Research*, 12, 1-18.
- Holmstrom, B. (1982) “Moral Hazard in Teams” *The Bell Journal of Economics*, Vol. 13, No. 2 (Autumn 1982), 324-340
- Holmstrom, B. and Milgrom, P. (1991) “ Multitask Principal Agent Analyses: Incentive Contracts, Asset Ownership and Job Design” *Journal of Law, Economics and Organisation*, 7, 24-52
- Ittner, Christopher D., David F. Larcker, and Marshall W. Meyer (2003) “Subjectivity and the Weighting of Performance Measures: Evidence From a Balanced Scorecard”, *Accounting Review*, 78 (3), 725-758
- Jensen, M. “Value Maximization, Stakeholder Theory, and the Corporate Objective Function” *Journal of Applied Corporate Finance*, Fall 2001
<http://papers.ssrn.com/abstract=220671>
- Kahneman, D. & Tversky, A. (1979) Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47, 263-291.
- Kandel, E and Lazear, E (1994) “Peer Pressure and Partnerships” *Journal of Political Economy*, 100 (4), 801-817
- Kaplan, R and D. Norton (1992) “The Balanced Scorecard - Measures that Drive Performance” *Harvard Business Review* January-February, 71-79.
- Kerr, S. (1995) “On the Folly of Rewarding A While Hoping for B”, *Academy of Management Executive*, 9 (1), 7-14.
- Knez, M and Simester, D (2001) “Firm-wide incentives and mutual monitoring at Continental Airlines” *Journal of Labour Economics*, 19 (4), 743-772
- Larkin, Ian I. (2006) “The Cost of High-powered Incentives: Salesperson Gaming in Enterprise Software” Harvard Business School Working Paper
- Lazear, E (2001) “Performance pay and productivity”, *American Economic Review*, 90: 5, 1346-1361
- MacInnis, D.J.; Moorman, C.M. and Jaworski, B.J. (1991) “Enhancing and Measuring Consumers’ Motivation, Opportunity and Ability to Process Brand Information from Ads,” *Journal of Marketing*, 55 (October), pp. 32–53.
- Maisel, L. (2001) “Performance measurement practices survey results”, AICPA, USA.
- Malina, M.A and Selto, F.H. (2001) “Communicating and Controlling Strategy: An Empirical Study of the Effectiveness of the Balanced Scorecard”, *Journal of Management Accounting Research*, 13, 47-90.
- Marr, B., Neely, A., Franco, M., Wilcox, M., Adams, C. and Manson, S. (2004), “Business Performance Measurement - What Is the State of the Art?”

- Proceedings of the Fourth International Performance Measurement Association Conference, Edinburgh, UK.
- Neely, A.D.; Kennerley, M. and Martinez, V. “Does the Balanced Scorecard Work: An Empirical Investigation”, Proceedings of the Fourth International Performance Measurement Association Conference, Edinburgh, UK.
- Neely, A.D., Yaghi, B. and Youell, N. (2008) “Enterprise Performance Management: The Global State of the Art”, Oracle and Cranfield School of Management.
- Nørreklit, H. (2000) “The Balance on the Balanced Scorecard: A Critical Analysis of Some of Its Assumptions”, *Management Accounting Research*, 11, 65-88.
- Nørreklit, H. (2003) “The Balanced Scorecard: What Is the Score? A Rhetorical Analysis of the Balanced Scorecard”, *Accounting, Organisations and Society*, 28, 591-619.
- Oyer, Paul (2004), “Why Do Firms Use Incentives that have No Incentive Effects?” *The Journal of Finance*, Volume 59, Number 4, August 2004 , pp. 1619-1650
- Parker, S.W. and Wall, T.D. (1998) “Job and Work Design: Organizing Work to Promote Well-Being and Effectiveness”, Sage Publications.
- Prendergast, C (1999) “The Provision of Incentives in Firms” *Journal of Economic Literature*, Vol. XXXVII (March 1999), 7-63
- Ridgway, V.F. (1956), “Dysfunctional Consequences of Performance Measurements”, *Administrative Science Quarterly*, 1(2), pp. 240-247.
- Rigby, D. (2001), “Management Tools and Techniques: A Survey”, *California Management Review* 43(2), 139-160.
- Speckbacher, G., Bischof, J. and Pfeiffer, T. (2003), “A Descriptive Analysis on the Implementation of Balanced Scorecards in German-speaking Countries”, *Management Accounting Research*, 14, pp. 361-387.
- Tversky, A. & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference Dependent Model. *Quarterly Journal of Economics* 106, 1039-1061.

Table 1: Balanced Scorecard measures

Financial measures

Return on Capital Employed	$[PBIT/(Debtors + Stock + Fixed Assets)] \times 100$
Growth in Profit	$[(Contribution\ This\ Year\ To\ Date - Contribution\ Last\ Year\ To\ Date) / Contribution\ Last\ Year\ To\ Date] \times 100$
PBIT as a % of Sales	$(Contribution\ YTD / Sales\ YTD) \times 100$
Positive Cash Flow	$[(Contribution - (+/-\ Stock\ Movement\ £'s) + (+/-\ Debtors\ Movement\ £'s) = Basic\ Cash\ Flow) / Total\ Sales] \times 100$
Sales Growth	$[(Sales\ PWD\ This\ Year\ To\ Date - Sales\ PWD\ Last\ Year\ To\ Date) / Sales\ PWD\ Last\ Year\ To\ Date] \times 100$

Customer measures

Customer Satisfaction ^b	<i>Score achieved via an external survey</i>
Customer Retention	$[(No.\ of\ Customers\ retained\ in\ rolling\ 12\ months\ to\ current\ month - No.\ of\ Customer\ retained\ in\ rolling\ 12\ months\ to\ last\ month) / No.\ of\ Customers\ retained\ in\ rolling\ 12\ months\ to\ last\ month] \times 100$
Sales Mix	$[(Sales\ of\ Selected\ SPGs\ This\ Year\ to\ Date - Sales\ of\ Selected\ SPGs\ Last\ Year\ to\ Date) / Sales\ of\ Selected\ LLSPGs\ Last\ Year\ to\ Date] \times 100$
Availability of Stock Range	$(Sum\ of\ Number\ of\ Days\ where\ Stock\ Ins\ for\ your\ MBR\ are\ equal\ to\ or\ greater\ than\ 90\% / Number\ of\ Trading\ Days) \times 100$

Internal measures

Operational Efficiency	<i>Stock/Debtors/Labour/Transport – Yes/No against individual targets: Stock 40 days, Debtors 0.5% against Sales, Labour 10% against Ex-Stock Sales, Transport 8% against Delivered Sales, where 25% is awarded per point</i>
Operational Standards	$(Score\ from\ Operational\ Standards\ Check\ List / Total\ possible\ score\ from\ Operational\ Standards) \times 100$
Inter-company Co-operation	$[(Number\ of\ Customers\ trading\ with\ foreign\ Branches\ This\ YTD - Number\ of\ Customers\ trading\ with\ foreign\ Branches\ Last\ YTD) / Number\ of\ Customers\ trading\ with\ foreign\ Branches\ LYTD] \times 100$

People measures

Staff retention	$(Number\ of\ voluntary\ leavers\ on\ a\ rolling\ 12\ month\ basis / Average\ head\ count\ in\ rolling\ 12\ months) \times 100$
Employee satisfaction	$(The\ number\ of\ people\ who\ indicate\ they\ are\ satisfied\ at\ work / average\ number\ of\ employees\ over\ the\ period) \times 100$
Communication	$(Number\ of\ people\ who\ feel\ they\ have\ been\ made\ aware\ of\ businesses\ activities / Average\ number\ of\ employees\ over\ the\ period) \times 100$ (By Region)

Supplier measures

Spend with Approved Suppliers	$(Purchases\ from\ preferred\ Suppliers\ This\ Year\ To\ Date / Total\ purchases\ from\ Suppliers\ This\ Year\ To\ Date) \times 100$
-------------------------------	--

Table 2: Mean (in £,000) and standard deviation for Division 1 branches and matched Division 2 branches

	Division 1	Division 2 (matched only)
<i>Branches</i>	<i>156</i>	<i>121</i>
<i>Observations</i>	<i>11076</i>	<i>11076</i>
Sales	152.5 (87.8)	166.4 (103.4)
Gross profits	45.7 (28.3)	39.9 (19.1)
Trading profits	15.3 (19.4)	22.8 (15.6)
Labour costs	14.9 (7.7)	8.8 (4.2)
Infrastructure	7.6 (4.7)	4.5 (2.4)
Transport costs	4.2 (2.6)	1.1 (1.1)
General and administration	1.4 (1.3)	0.8 (0.6)
Other	0.9 (0.7)	0.7 (0.5)
IT	0.48 (0.22)	0.49 (0.13)
Marketing and advertising	0.15 (0.34)	0.02 (0.15)

Notes: Values are monthly in nominal £,000 over the period August 1999 to July 2005.

Table 3: Descriptive statistics, weather and demand

	Mean (standard deviation)
Minimum temperature (in Celsius), measured at 26 points throughout the UK	7.28 (4.17)
Rain fall (in mm)	65.7 (41.0)
National quarterly demand for activities using Division 1 products (in £m)	2628 (200)
National quarterly demand for activities using Division2 products (in £m)	2588 (168)

Notes: Data on weather if from <http://www.met-office.gov.uk/climate/uk/stationdata/>

Table 4: Job type, number of employees and average experience

Job	Mean number per branch	Std. Dev.	Min	Max	Mean experience	Mean age
Manager	0.883	0.359	0	2	12.971 (9.876)	42.088 (8.570)
Office	1.090	0.808	0	3	8.821 (9.062)	41.240 (11.825)
Sales	2.548	1.667	0	9	7.088 (8.673)	38.532 (12.804)
Driver	1.651	1.209	0	6	6.520 (7.747)	46.443 (10.576)
Warehouse and stockyard	3.670	2.107	0	14	5.771 (7.666)	38.970 (13.330)
Other	0.354	0.543	0	2	7.579 (7.476)	47.572 (13.564)
<i>Total</i>	<i>10.2</i>	<i>4.631</i>	<i>3</i>	<i>30</i>	<i>7.234</i> <i>(8.520)</i>	<i>40.883</i> <i>(12.637)</i>

Manager: Branch Manager, Branch Manager (Designate) ; **Office:** Administration Assistant, Administration Supervisor, Administrator, Assistant Branch Manager, Assistant Depot Manager, Branch Supervisor , Deputy Manager, Estimator, General Clerk, Typist, Office Manager, Secretary ; **Sales:** Contracts/Sales Administrator, Credit Controller, Goods Inwards Assistant, Inside Sales, Sales Supervisor, Sales Clerk, Sales Assistant, Sales Negotiator, Sales Representative, Showroom Supervisor, Stock Controller ; **Driver:** Driver, Glass Cutter/Driver, Warehouse Assistant/Driver; **Warehouse/Yard:** Depot Manager, Depot Manager Designate, Drainage Supervisor, Foreman, Heavy Supervisor, Hire Assistant, Ironmongery Supervisor, Logistics Manager, Maintenance Supervisor, Operations Manager, Product Supervisor, Stores Assistant, Warehouse Assistant, Yard Assistant, Saturday Assistant, Timber Supervisor, Transport Supervisor, Warehouse Supervisor, Yard Manager, Yard Supervisor ; **Other:** Fixer, Machine Operator, Mill Operative, Cleaner, Timber Machinist, Trainee.

Table 5: Matched/Diff-in-Diff results on sales, different time periods

	(1)	(2)	(3)	(4)
Dep var: difference in Sales				
Nov 2001 - Jul 2002				404 (2382)
Aug 2002 - Jul 2004	4538 (1786)**			
Aug 2002 - Jul 2003		3752 (2624)	8305 (2435)***	8671 (3577)**
Aug 2003 - Jul 2004		4992 (2136)**	8903 (4020)***	9216 (4625)**
Aug 2004 - Jul 2005			6916 (4872)	7124 (5157)
Demand	-25 (28)	-23 (30)	-50 (24)**	-52 (20)***
Min temp in C	1686 (455)***	1694 (463)***	1354 (453)***	1348 (455)***
Rain fall in mm	-90 (16)***	-89 (15)***	-80 (13)***	-79 (13)***
R-squared	0.15	0.15	0.15	0.15

Note: Standard errors in parentheses are clustered at the branch level. 11,076 observations on 156 branches over the period August 1999 to July 2005. Constant, month and branch dummies included in all regressions. Dependent variable is the difference in sales in a division 1 branch and the geographically nearest division 2 branch. Demand is the difference in national demand for division 1 products and demand for division 2 products.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: Matched/Diff-in-Diff results on all variables

Dep var: difference in	(1) Sales	(2) Gross profits	(3) Trading profits	(4) Labour costs	(5) Infrastructure	(6) Transport costs	(7) General and administration	(8) Other	(9) IT	(10) Marketing and advertising
Aug 2002 - Jul 2003	8305 (2435)***	1512 (591)**	771 (677)	192 (161)	91 (125)	704 (82)***	28 (45)	-12 (37)	-59 (8)***	12 (14)
Aug 2003 - Jul 2004	8903 (4020)**	2565 (976)***	995 (846)	675 (261)**	183 (211)	1014 (112)***	115 (49)**	-48 (42)	-78 (10)***	19 (18)
Aug 2004 - Jul 2005	6916 (4872)	1835 (1155)	-1018 (932)	933 (299)***	667 (309)**	1154 (119)***	88 (50)*	-72 (37)*	-91 (10)***	-4 (10)
Demand	-49 (23.604)* *	-4.08 (5.331)	-7.61 (5.087)	-2.161 (1.426)	1.732 (1.007)*	-0.594 (-0.669)	0.200 (0.325)	-0.577 (0.280)**	-0.333 (0.070)***	-0.163 (0.094)*
Min temp in C	1354 (452.768) ***	527 (135.379)***	624 (157.243)***	-6.60 (30.43)	-88.9 (55.781)	-40.6 (18.040)**	34.9 (10.624)***	-4.593 (6.902)	-5.457 (2.418)**	0.372 (2.584)
Rain fall in mm	-79 (13.352)* **	-19.0 (3.389)***	-22.8 (4.107)***	0.144 (0.735)	-0.957 (0.955)	0.167 (0.447)	-0.724 (0.378)*	0.840 (0.255)***	0.017 (0.054)	0.100 (0.101)
R-squared	0.15	0.20	0.13	0.05	0.01	0.08	0.01	0.01	0.05	0.00

Note: Standard errors in parentheses are clustered at the branch level. 11,076 observations on 156 branches over the period August 1999 to July 2005. Constant, month and branch dummies included in all regressions. Dependent variable is the difference in variable indicated in a Division 1 branch and the geographically nearest Division 2 branch. Demand is the difference in national demand for Division 1 products and demand for Division 2 products.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: Matched/Diff-in-Diff results on all variables - comparison of branches on non-financial balanced scorecard measures

branch score on non-financial balanced scorecard measures Dep var: difference in	(1)	(2)	(3)	(4)
	below median		above median	
	Sales	Trading profits	Sales	Trading profits
Aug 2002 - Jul 2003	1947 (3740)	-838 (1062)	14676 (2964)***	475 (167)***
Aug 2003 - Jul 2004	3535 (6135)	-979 (1260)	14266 (5173)***	822 (230)***
Aug 2004 - Jul 2005	3 (7538)	-2725 (1437)*	13583 (6097)**	1182 (278)***
Demand	-44.59 (31.829)	-7.457 (7.277)	-55.526 (35.113)	-2.122 -2.133
Min temp in C	1063.194 (671.935)	631.729 (216.424)***	1676.383 (568.785)***	-7.446 -38.592
Rain fall in mm	-48.264 (18.298)**	-19.621 (5.172)***	-113.104 (18.036)***	-0.652 -0.977
R-squared	0.12	0.09	0.21	0.08

Note: Standard errors in parentheses are clustered at the branch level. Period August 1999 to July 2005. Constant, month and branch dummies, demand, min temperature and monthly rainfall included in all regressions. Dependent variable is the difference in variable indicated in a Division 1 branch and the geographically nearest Division 2 branch. Demand is the difference in national demand for Division 1 products and demand for Division 2 products.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8: Matched/Diff-in-Diff results on all variables - comparison of branches on non-financial balanced scorecard measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
branch score on financial measures pre-BSC	below median				above median			
branch score on non-financial BSC measures	below median		above median		below median		above median	
Dep var: difference in	Sales	Trading profits	Sales	Trading profits	Sales	Trading profits	Sales	Trading profits
Aug 2002 - Jul 2003	1665 (5125)	-418 (1190)	19833 (5786)***	2605 (1259)**	2275 (5396)	-1428 (2017)	11454 (3152)***	2265 (1070)**
Aug 2003 - Jul 2004	678 (8076)	-147 (1475)	20045 (10554)*	2626 -1706	7908 (9544)	-2291 (2293)	10631 (5238)**	3227 (1429)**
Aug 2004 - Jul 2005	-7080 (10227)	-2269 (1837)	16018 (10514)	-546 (1389)	10617 (10827)	-3601 (2335)	11981 (7505)	1444 (1663)
Demand	-53.493 (33.908)	-3.43 (7.265)	-139.816 (58.944)**	-10.066 -10.074	-31.101 -62.969	-14.723 (14.902)	-3.042 (42.45)	-6.414 (9.887)
Min temp in C	1098.729 (876.553)	281.484 (252.039)	457.288 -928.409	34.444 -283.457	1483.235 -1014.43	1303.259 (364.396)***	2484.841 (726.715)***	964.148 (298.059)***
Rain fall in mm	-19.043 (20.931)	-10.703 (5.808)*	-117.575 (32.175)***	-25.176 (11.438)**	-112.138 (36.930)***	-39.426 (11.212)***	-112.307 (22.354)***	-24.764 (7.177)***
R-squared	0.1	0.06	0.16	0.14	0.17	0.2	0.27	0.22

Note: Standard errors in parentheses are clustered at the branch level. Period August 1999 to July 2005. Constant, month and branch dummies, demand, min temperature and monthly rainfall included in all regressions. Dependent variable is the difference in variable indicated in a Division 1 branch and the geographically nearest Division 2 branch. Demand is the difference in national demand for Division 1 products and demand for Division 2 products.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 9: Years of service

Dep var: difference in	(1)	(2)	(3)	(4)
	Sales		Trading profit	
<i>[interaction is evaluated at the mean level of service = 6.6 years]</i>				
service x (Aug02-Jul03)	1724 (909)*	798 (1014)	401 (203)*	110 (232)
service x (Aug03-Jul04)	3181 (1229)**	2458 (1413)*	659 (220)***	192 (262)
service x (Aug04-Jul05)	3857 (1479)**	3026 (1662)*	459 (222)**	48 (270)
<i>[interaction is evaluated at the mean level of service = 11.3 years]</i>				
senior service x (Aug02-Jul03)		589 (388)		185 (69)***
senior service x (Aug03-Jul04)		460 (529)		297 (92)***
senior service x (Aug04-Jul05)		528 (618)		261 (117)**
Aug 2002 - Jul 2003	8320 (2428)***	8317 (2487)***	771 (679)	770 (675)
Aug 2003 - Jul 2004	8939 (3965)**	8932 (3989)**	998 (835)	994 (818.)
Aug 2004 - Jul 2005	6938 (4770)	6932 (4772)	-1016 (921)	-1017 (904)
R2	0.16	0.17	0.13	0.13

Note: Standard errors in parentheses are clustered at the branch level. 11,076 observations on 156 branches over the period August 1999 to July 2005. Constant, month and branch dummies, demand, min temperature and monthly rainfall included in all regressions. Dependent variable is the difference in variable indicated in a Division 1 branch and the geographically nearest Division 2 branch. Demand is the difference in national demand for Division 1 products and demand for Division 2 products.

* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 1: The Scorecard measures

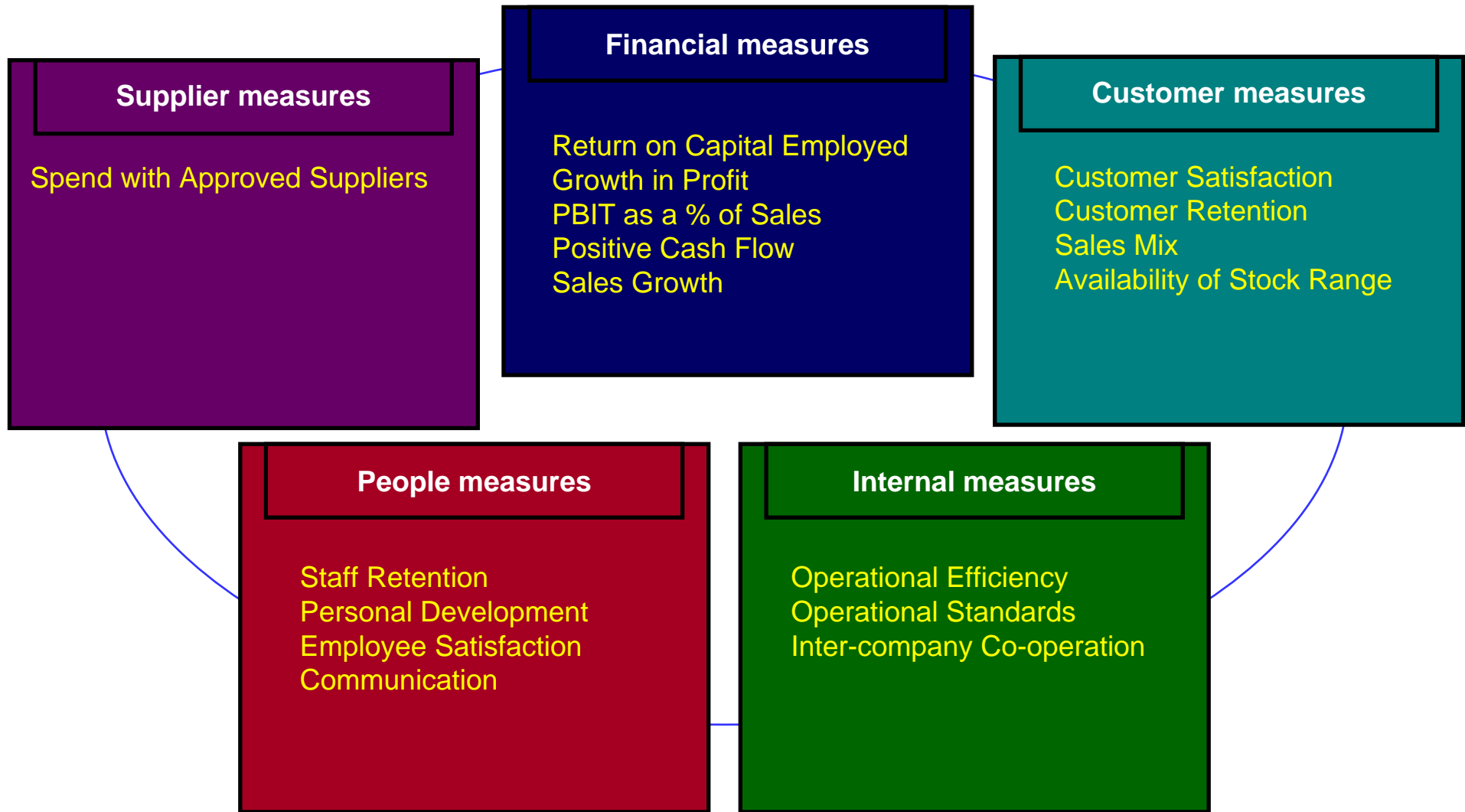


Figure 2: Sales in all Division 1 branches relative to Division 2 branches

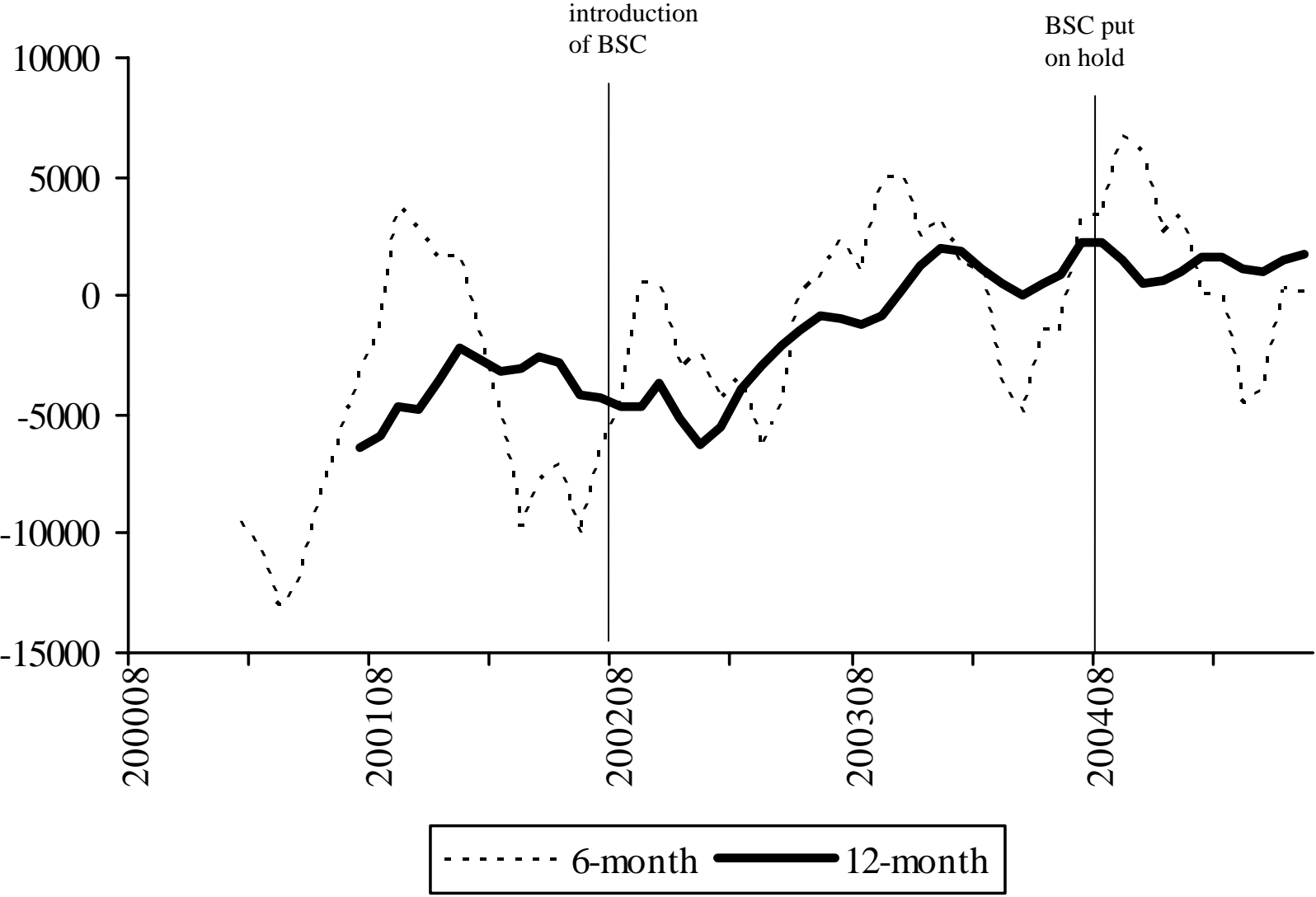


Figure 3: Trading profits in all Division 1 branches relative to Division 2 branches

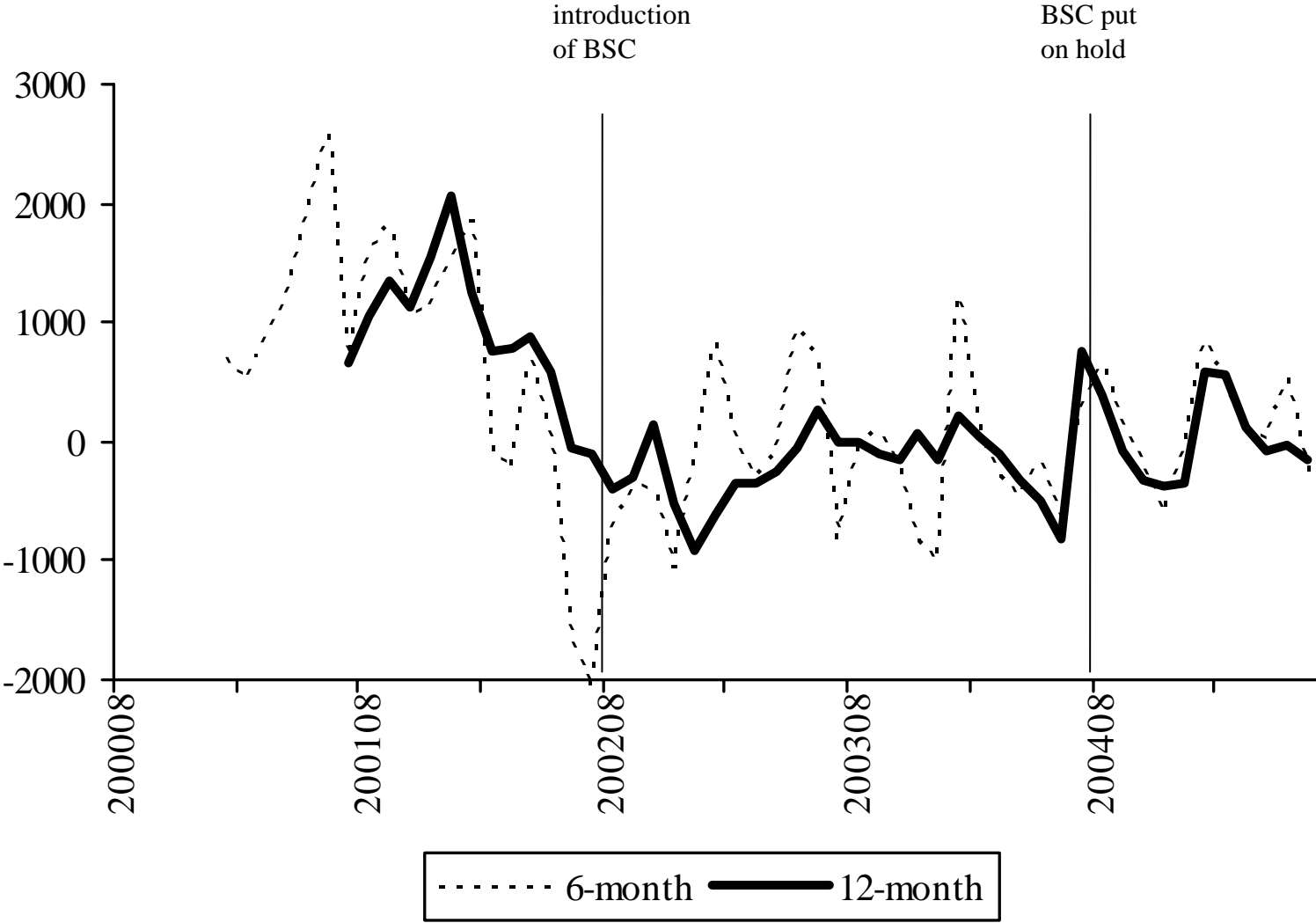
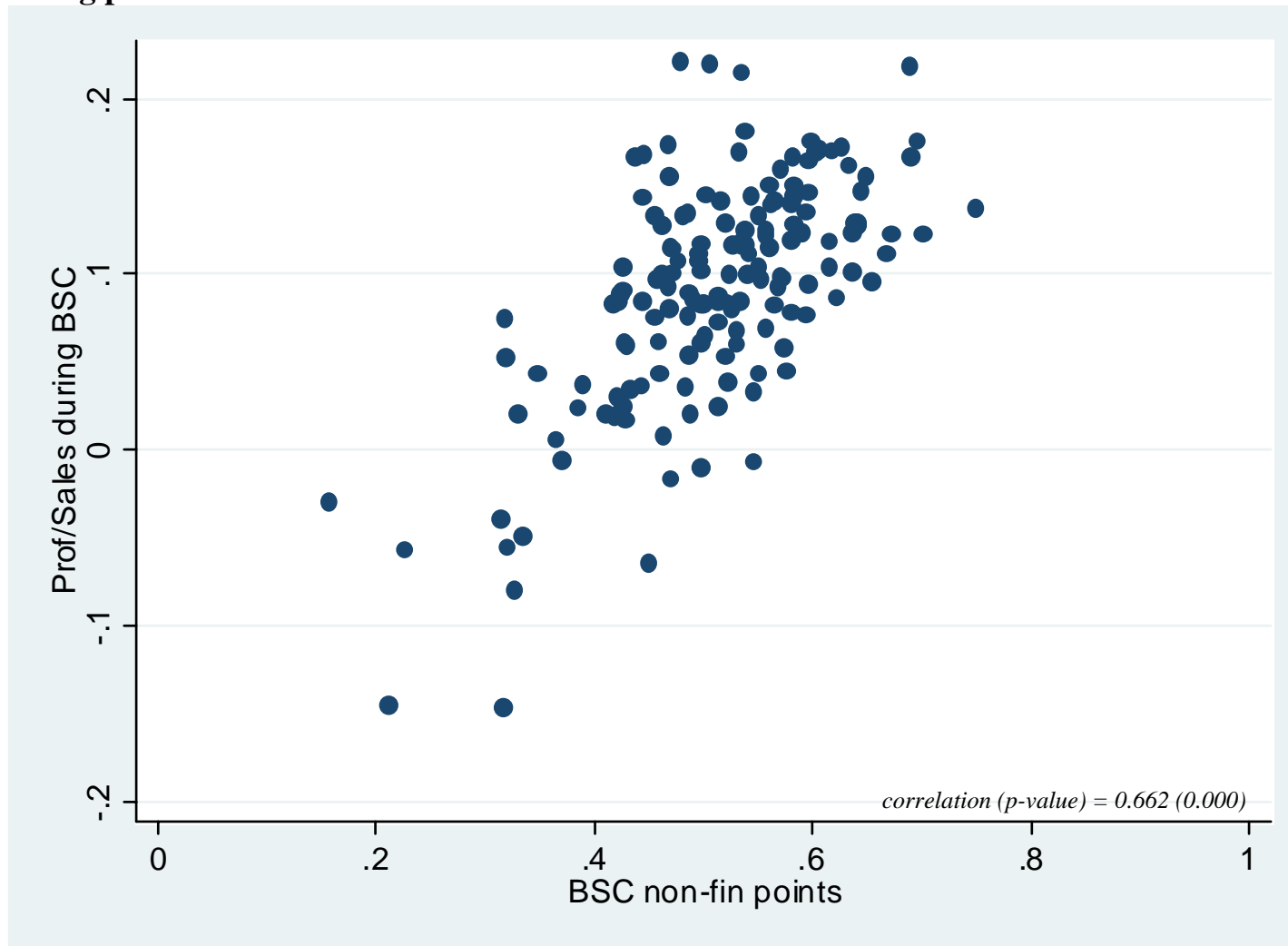
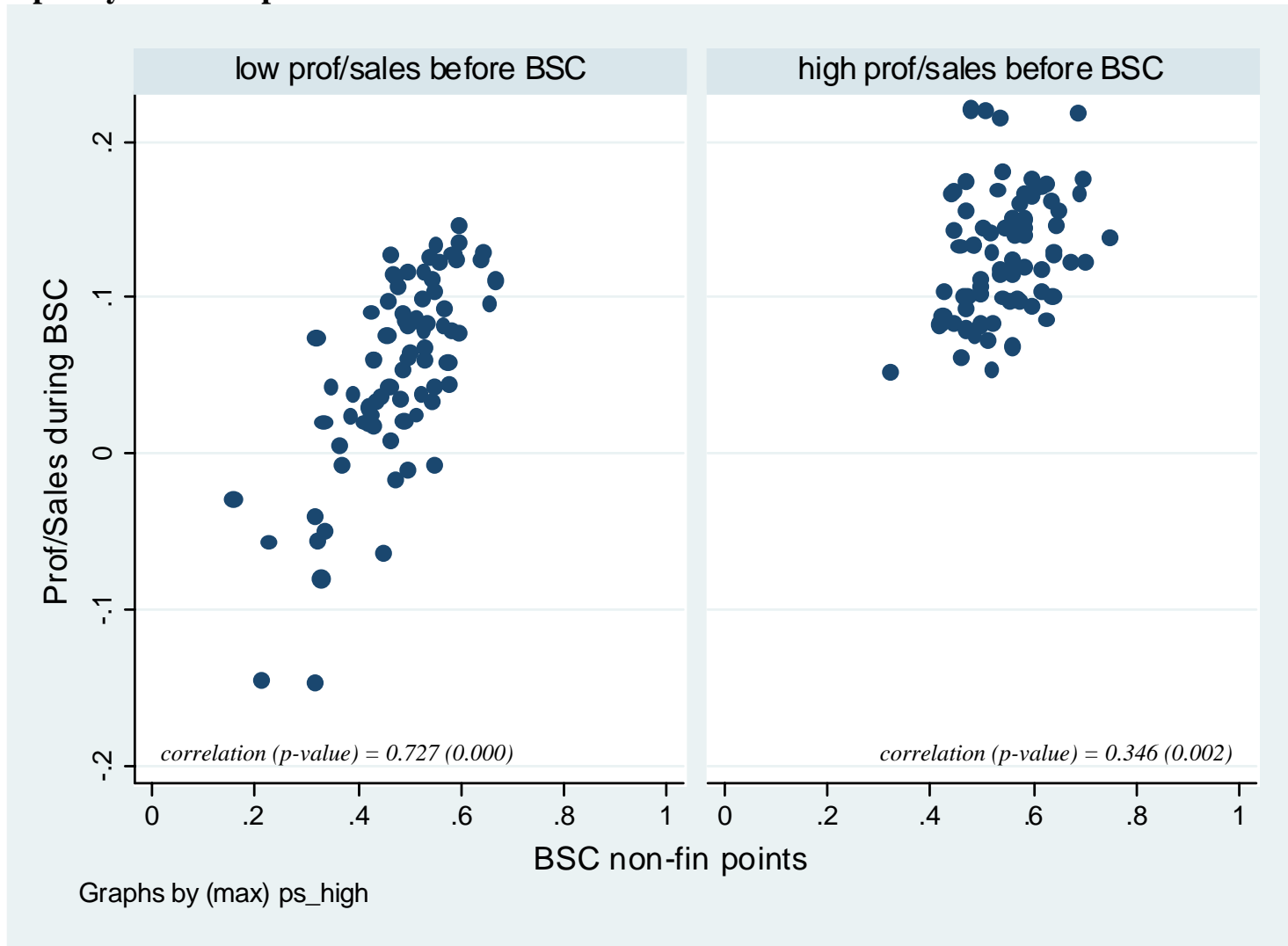


Figure 4: Comparison of non-financial with financial performance during Balanced Scorecard, using profits over sales rather than BSC measures



Notes: 156 observations (one for each Division 1 Brand A branch); x-axis is the share that each branch earned of the total points they could have earned on the non-financial BSC measures (customer, internal, people and supplier) over the period August 2002 - July 2004; y-axis is the average value of trading profits over sales for each branch over the period August 2002 - July 2004.

Figure 5: Comparison of non-financial with financial performance during Balanced Scorecard, split by financial performance before BSC



Notes: 156 observations (one for each Division 1 Brand A branch); x-axis is the share that each branch earned of the total points they could have earned on the non-financial BSC measures (customer, internal, people and supplier) over the period August 2002 - July 2004; y-axis is the average value of trading profits over sales for each branch over the period August 2002 - July 2004; left-hand panel are those branches that had below median ratio of profit to sales over the period August 1999 - July 2002.

Figure 6: Sales in all Division 1 branches relative to Division 2 branches, split by level of experience of senior staff

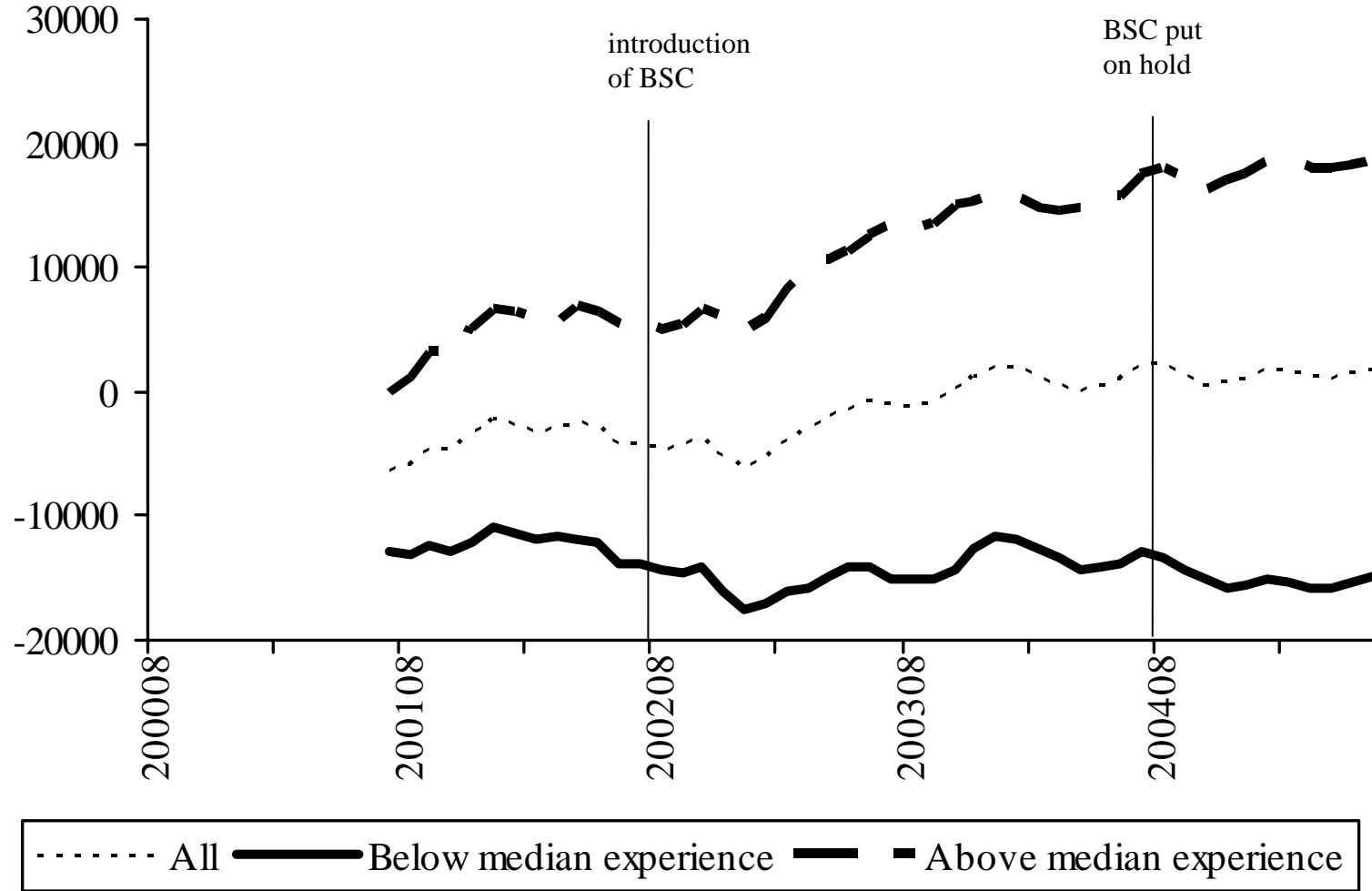


Figure 7: Trading Profit in all Division 1 branches relative to Division 2 branches, split by level of experience of senior staff

