

# DISCUSSION PAPER SERIES

No. 6107

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

Rachel Griffith and Andrew Neely

***LABOUR ECONOMICS***



**Centre for Economic Policy Research**

**[www.cepr.org](http://www.cepr.org)**

Available online at:

**[www.cepr.org/pubs/dps/DP6107.asp](http://www.cepr.org/pubs/dps/DP6107.asp)**

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

**Rachel Griffith**, Institute for Fiscal Studies, University College London and CEPR  
**Andrew Neely**, Cranfield School of Management

Discussion Paper No. 6107  
February 2007

Centre for Economic Policy Research  
90–98 Goswell Rd, London EC1V 7RR, UK  
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999  
Email: [cepr@cepr.org](mailto:cepr@cepr.org), Website: [www.cepr.org](http://www.cepr.org)

This Discussion Paper is issued under the auspices of the Centre's research programme in **LABOUR ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Rachel Susan Griffith and Andrew Neely

CEPR Discussion Paper No. 6107

February 2007

## ABSTRACT

### Incentives and Managerial Experience in Multi-Task Teams: Evidence from Within a Firm\*

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.

JEL Classification: J33, M12 and M52

Keywords: balanced scorecard, incentive design and managerial experience

Rachel Griffith  
Institute for Fiscal Studies  
7 Ridgmount Street  
London  
WC1E 7AE  
Email: rgriffith@ifs.org.uk

Andrew Neely  
Cranfield University  
School of Management  
Bedfordshire MK43 OAL  
United Kingdom  
Email: a.neely@cranfield.ac.uk

For further Discussion Papers by this author see:  
[www.cepr.org/pubs/new-dps/dplist.asp?authorid=125339](http://www.cepr.org/pubs/new-dps/dplist.asp?authorid=125339)

For further Discussion Papers by this author see:  
[www.cepr.org/pubs/new-dps/dplist.asp?authorid=166014](http://www.cepr.org/pubs/new-dps/dplist.asp?authorid=166014)

\* 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). All errors and omissions remain the responsibility of the authors.

Submitted 10 January 2007

# 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 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, 1991). Holmstrom and Milgrom also suggest that tasks should be grouped together, based on the cost of measuring and rewarding performance in that dimension. 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 incentives 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 a large number of firms and organisations across the globe.<sup>1</sup> The idea behind the Balanced Scorecard is that managers can improve performance by monitoring and rewarding a large range of linked activities (usually 15-20 covering four different dimensions – financial, customer, internal and innovation/learning). These activities are designed to reflect both current actions that are assumed to 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 versus all of the indicators. Thus as well as forming the basis for

---

<sup>1</sup> See, for example, the Balanced Scorecard Hall of Fame at [http://www.bscol.com/pdf/BSCHoF-Membersby\\_Industry\\_2000-2005.pdf](http://www.bscol.com/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

an incentive scheme, the Balanced 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.<sup>2</sup> Some of the assumptions of causality that underpin the designed of balanced scorecards are challenged (Nørreklit, 2000; Nørreklit, 2003). Ittner et al. (1998) present a compelling case illustrating the difficulty of basing an incentive scheme on a balanced scorecard. While 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 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 empirical evidence on how well the scheme works.<sup>3</sup>

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 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

---

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.

<sup>2</sup> See, inter alia, Cools and van Praag (2003), Fink (2004), Gosling (2003), Jensen (1991, 2001) and Ittner, Larcker and Meyer (1998), Nørreklit (2000), Nørreklit (2003). There is also an earlier literature, for example, Ridgway (1956) criticises the idea of multidimensional incentives schemes.

<sup>3</sup> Hoque and James (2000) survey 66 Australian manufacturing firms, Banker, Potter and Srinivsan (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.

- by comparing monthly performance across a large number of branches in two different divisions.<sup>4</sup>

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, rather than conflate the impact of any form of incentive pay with this particular form of incentive pay. We also investigate the idea that it was not the incentive aspect of the Balanced Scorecard that was influential, but that it improved the information that was transmitted to managers and workers and this helped them to focus their actions on more profitable margins.

We find that the Balanced Scorecard changed behaviour, but that overall this change did not lead to increased profits in all the branches where it was implemented – at the aggregate level costs increased by at least as much as sales at the brand level. However, the impact varied across branches - branches with more experienced managers were able to effectively improve performance, while less experienced managers were not able to effectively improve performance. We use interviews with a number of individual managers to explore the idea that the impact of experience enabled managers to better interpret the large number of indicators and to effectively allocate effort within the branch.

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

---

<sup>4</sup> 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.

Balanced Scorecard as showing that it requires additional ability (acquired through experience) for the manager to know where the greatest pay off will be. We corroborate this interpretation in interviews with a large number of branch managers.

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 design 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 product (which are both used in building). Another distinction is that the average value of and margin earned on the product sold in Division 2 is higher. However, the branches operate in similar economic conditions, for example, they employ people from the same labour markets, and experience similar demand shocks.

The dominant component of firm profits is individual branch profits, though there are also orders received at head office and volume discounts that head office receives from suppliers (called rebates). 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,<sup>5</sup> 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.

---

<sup>5</sup> 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.

## **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 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.<sup>6</sup> 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. 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).<sup>7</sup>

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-

---

<sup>6</sup> See, inter alia, review by Prendergast (1999) and Hayes and Abernathy (1980).

<sup>7</sup> See Argyris (1952), Ridgway (1956), Holmstrom and Milgrom (1991), Kerr (1995) and for a recent review of the literature Prendergast (1999).

monthly basis, but effectively earned on a monthly basis. For branch staff and branch managers the payment was based on the number of points the branch earned in the month times the value of a point.<sup>8</sup> 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. The average salary of branch workers is around £12,000, so the maximum bonus represents 5.1% of their salary.

The total amount the firm allocated for the scheme was £1,836,000. 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.<sup>9</sup> 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 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,

---

<sup>8</sup> 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.

<sup>9</sup> The firm estimates the cost of implementation at around £0.5m, including the direct and indirect costs such as management time.

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 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 incentive 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.<sup>10</sup> In the end the firm adopted 17 measures, shown in Figure 2 and described further in Tables 1.<sup>11</sup>

[Figure 1 around here]

[Table 1 around here]

Following a 9 month design and deployment process the Balanced Scorecard was introduced in business from 1<sup>st</sup> August 2002. The business' financial year ran from 1<sup>st</sup> August-31<sup>st</sup> July and the scheme was operated for the time period 1<sup>st</sup> August 2002-31<sup>st</sup> December 2005. For the purposes of this evaluation we will focus on two financial years 1<sup>st</sup> August 2002-31<sup>st</sup> July 2003 and 1<sup>st</sup> August 2003-31<sup>st</sup> July 2004.

---

<sup>10</sup> 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.

## **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 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 changes. 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

---

<sup>11</sup> 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.

improvements in performance - then workers will devote more time to those tasks that are measured more precisely.<sup>12</sup> 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 was very aggregated, hence managers were not able to directly take action based on them. Each manager need 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 then this time last year is not sufficient information to enable the manager or shop workers to be able to act on it. They needed to be able to figure out which customers were not coming back and why in order to know where to put effort in order to increase retention rates. This is where experience comes in. Our thesis 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 credible) motivate staff to put effort into the activities that matter for performance; with an inexperience managers workers can easily become overloaded with information and

---

<sup>12</sup> See Holmstrom and Milgrom (1991) and Burgess et al (2004).

tasks, leading to underperformance. We explore this idea in our empirical analysis below.

### **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 employees 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 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 one are used both inside and outside, 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. Tables 4 and 5 show 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 and 5 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 brand A branches which do not have a Division 2 brand C 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 method<sup>13</sup> and use a combination of matching and a 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. We control for the factors that affect the branches differently, such as the weather.

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

---

<sup>13</sup> See, inter alia, Cook and Campbell (1979) and Heckman, Ichimura and Todd (1997). For a recent survey see Blundell and Costa-Dias (2005).

$$(3) \quad \begin{aligned} S_{irt} &= \alpha_i + \beta_1 L_{rt} + \beta_2 D_{1t} + \beta_3 \tau_t + \beta_4 W_{rt} + \gamma BSC_t + e_{irt} \\ S_{jrt} &= \alpha_j + \lambda_1 L_{rt} + \lambda_2 D_{2t} + \lambda_3 \tau_t + \lambda_4 W_{rt} + \varepsilon_{jrt} \end{aligned}$$

where S: sales, L: local economic factors such as labour markets, supply chain, etc., D: national demand,  $\tau$  : seasonal (monthly) dummies, W: weather, BSC: indicators of the use of the balanced scorecard,  $e, \varepsilon$  : idiosyncratic shocks.

We match each branch from brand A to the geographically nearest branch from brand C and consider the difference between the two braches. This takes the form

$$(4) \quad \begin{aligned} (S_{irt} - S_{jrt}) &= (\alpha_i - \alpha_j) + (\beta_1 - \lambda_1)L_{rt} + (\beta_2 D_{1t} - \lambda_2 D_{2t}) \\ &+ (\beta_3 - \lambda_3)\tau_t + (\beta_4 - \lambda_4)W_{rt} + \gamma BSC_t + (e_{irt} - \varepsilon_{jrt}). \end{aligned}$$

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:

$$(4') \quad \begin{aligned} (S_{irt} - S_{jrt}) &= (\alpha_i - \alpha_j) + \beta_2 (D_{1t} - D_{2t}) + (\beta_3 - \lambda_3)\tau_t \\ &+ (\beta_4 - \lambda_4)W_{rt} + \gamma BSC_t + (e_{irt} - \varepsilon_{jrt}). \end{aligned}$$

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 managerial experience,  $\gamma = \gamma_0 + \gamma_1 Experience_i$ .

## 4 Results

### 4.1 The overall impact of the balanced scorecard

We start in Table 6 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 brand A branches with brand C branches during this period with the difference in the

level before and after this period. This shows that sales in brand A branches increased by an average of £4,538 per month more than brand C 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. 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 just to the two years before the Balanced Scorecard was implemented. We now see a larger impact around (£8,305) which diminishes slightly over time. In column (4) we consider whether there was an anticipation effect in the initial discussion period (i.e. a Hawthorne effect in anticipation of the actual implementation), but find no evidence of this. For the remainder of this paper we stick to the parameterisation in column (3). In Table 6 we also see that demand for brand C products grew somewhat faster than brand A products (the coefficient on demand is negative), that there is more demand for brand A products when it is hotter, and less when it is raining (this makes sense as brand is more often used outside).

[Table 6 around here]

These results suggest that the Balanced Scorecard had an impact on increasing sales, but what about profits? In Table 7 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 effected.

[Table 7 around here]

The first column of Table 7 repeats column (3) of Table 6. 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.

#### **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 8 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 top half of Table 8 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 in the bottom half those branches that did well on the non-financial measures.

[Table 8 around here]

Here we see large differences between the two groups. In the top 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, gross profits and trading profits. While costs did increase in these branches, sales increased by more, resulting in higher profits.

Figure 2 shows a similar picture. Each dot represents a brand 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 2 around here]

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 during three time periods - prior to the Balance Scorecard (August 1999 - July 2002), during the Balanced Scorecard (August 2002 - July 2004) and after (August 2004 - July 2005).

[Figure 3 around here]

In Figure 3 we show that good performance on the non-financial points during the Balanced Scorecard also led to good financial performance afterwards. In Figure 4 we split by financial performance in the period before the Balanced Scorecard was introduced. 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).

[Figure 4 around here]

In Figure 5 we show this same picture separately for branches that did poorly in financial terms in the period before the Balanced Scorecard and those that did well, and we see a positive effect in both cases.

[Figure 5 around here]

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 of the form of equation (4) 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.

The top panel of Table 9a shows the estimates for group (1), the bottom panel for group (2), the top panel of Table 8b shows the estimates for group (3) and the bottom panel for group (4).

[Tables 9a and 9b around here]

What we see is a pattern that suggests that branches that do well on the non-financial measures also do well on financial measures. In particular, those branches that previously did badly on financial measures (bottom panel of Table 9a) had substantially higher sales during implementation period and less so after. While costs increased, sales increased by more, so that trading profits also increased.

Consider the branches that did well financially prior to implementation. Those that did badly on non-financial measures (top panel of Table 9b) did badly in financial terms after implementation, while those that did well on non-financial measures (bottom panel of Table 9b) did well in financial terms after implementation.

The second empirical result of the paper is to show 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***

We were interested in why the response of branch managers to the Balanced Scorecard seemed to differ markedly. 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].

We found from these that some managers found it very difficult to understand and use the information presented in the Balanced Scorecard report, e.g. responses included *"The honest truth is, probably 60% of the measurements I could understand"* and *"I ignored it completely"*. In contrast to this, some managers found it very useful: *"you could have a quick look and a quick snapshot of what was going on and if you looked at it harder you could find things... and it gave you the benefit to change things before you drifted off and got too bad"*

We noticed that it seemed to be more experienced managers that were able to utilise the information in the Balanced Scorecard, while newer managers found it more difficult. 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.

Table 10 repeats the results in Table 7 but allowing the impact of the Balanced Scorecard to vary with the average years of experience of all staff. In the first three columns we see that experience matters. A branch with staff with the average years of experience (6.6 years) will have a £400 higher level of trading profits after the Balanced Scorecard was introduced than a branch with all new staff. In the second three columns we 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 around £200 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.<sup>14</sup>

## 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 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.

---

<sup>14</sup> 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).

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

- 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.
- 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>
- 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/cmpos/research/incentives/incentablehk.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
- Lazear, E (2001) “Performance pay and productivity”, *American Economic Review*, 90: 5, 1346-1361
- 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.
- 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.
- 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 <sup>b</sup>	<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	$[(Sales\ of\ Selected\ SPGs\ This\ Year\ to\ Date - Sales\ of\ Selected\ LLSPGs\ Last\ Year\ to\ Date) / Sales\ of\ Selected\ LLSPGs\ Last\ Year\ to\ Date] \times 100$
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 Brand A	Division 2 Brand C (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 Brand A products (in £m)	2628 (200)
National quarterly demand for activities using Brand products (in £m)	2588 (168)

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

**Table 4: Descriptive Statistics, level of sales**

Dep Var: sales	(1)	(2)	(3)	(4)	(5)	(6)
	Division 1 Brand A			Division 2 Brand C (matched only)		
Aug 2002 - Jul 2003			4958 (1571)***			8276 (1705)***
Aug 2003 - Jul 2004			6986 (2535)***			14239 (3483)***
Aug 2004 - Jul 2005			10970 (2645)***			20639 (4258)***
Demand - Div 1	38 (4)***	40 (2)***	20 (6)***			
Demand - Div 2				35 (7)***	31 (3)***	-14 (10)
Min temp in C	3295 (543)***	778 (294)***	738 (294)**	-6903 (766)***	-604 (435)	-601 (437)
Rain fall in mm	-107 (22)***	-93 (8)***	-89 (8)***	-18 (30)	-11 (12)	-0 (12)
Observations	11,076	11,076	11,076	8,591	8,591	8,591
Branches	156	156	156	121	121	121
R-squared	0.06	0.34	0.35	0.03	0.15	0.15
Branch fixed effects		yes	yes		yes	yes

Note: Robust standard errors in parentheses. Period covered is August 1999 to July 2005. Constant and month dummies included in all regressions.

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

**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 (1727)
Aug 2002 - Jul 2004	4538 (1433)***			
Aug 2002 - Jul 2003		3752 (1670)**	8305 (1815)***	8671 (2520)***
Aug 2003 - Jul 2004		4992 (1557)***	8903 (1694)***	9216 (2299)***
Aug 2004 - Jul 2005			6916 (1525)***	7124 (1843)***
Demand	-25 (12)**	-23 (12)*	-50 (13)***	-52 (17)***
Min temp in C	1686 (446)***	1694 (446)***	1354 (451)***	1348 (453)***
Rain fall in mm	-90 (12)***	-89 (12)***	-80 (12)***	-79 (12)***
R-squared	0.15	0.15	0.15	0.15

Note: Robust standard errors in parentheses. 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**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep var: difference in	Sales	Gross profits	Trading profits	Labour costs	Infrastructure	Transport costs	General and administration	Other	IT	Marketing and advertising
Aug 2002 - Jul 2003	8305 (1815)***	1512 (548)***	771 (664)	192 (114)*	91 (120)	704 (73)***	28 (54)	-12 (32)	-59 (8)***	12 (14)
Aug 2003 - Jul 2004	8903 (1694)***	2565 (560)***	995 (619)	675 (104)***	183 (122)	1014 (58)***	115 (50)**	-48 (30)	-78 (8)***	19 (22)
Aug 2004 - Jul 2005	6916 (1525)***	1835 (450)***	-1018 (517)**	933 (93)***	667 (126)***	1154 (50)***	88 (52)*	-72 (25)***	-91 (8)***	-4 (15)
Demand	-49 (12)***	-4.08 (3.58)	-7.61 (4.17)*	-2.161 (0.785)***	1.732 (0.809)**	-0.594 (0.471)	0.200 (0.382)	-0.577 (0.218)***	-0.333 (0.067)***	-0.163 (0.108)
Min temp in C	1354 (451)***	527 (123)***	624 (148)***	-6.60 (27.95)	-88.9 (33.2)***	-40.6 (19.9)**	34.9 (11.1)***	-4.593 (8.180)	-5.457 (2.301)**	0.372 (2.821)
Rain fall in mm	-79 (12)***	-19.0 (3.3)***	-22.8 (3.9)***	0.144 (0.774)	-0.957 (0.807)	0.167 (0.484)	-0.724 (0.333)**	0.840 (0.255)***	0.017 (0.066)	0.100 (0.095)
R-squared	0.15	0.20	0.13	0.05	0.01	0.08	0.01	0.01	0.05	0.00

Note: Robust standard errors in parentheses. 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 brand A branch and the geographically nearest division 2 brand C 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**

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
<b>5,538 observations on 78 branches scoring below median on non-financial balanced scorecard measures</b>										
Aug 2002 - Jul 2003	1947 (2706)	-450 (772)	-838 (955)	-91 (169)	-64 (177)	667 (102)***	18 (77)	-2 (47)	-54 (11)***	17 (20)
Aug 2003 - Jul 2004	3535 (2508)	-40 (790)	-979 (871)	528 (160)***	-212 (173)	970 (94)***	98 (68)	-5 (44)	-79 (11)***	40 (27)
Aug 2004 - Jul 2005	3 (2328)	-353 (690)	-2725 (785)***	681 (147)***	490 (216)**	1121 (80)***	147 (89)*	-62 (37)*	-83 (13)***	-13 (20)
<b>5,538 observations on 78 branches scoring above median on non-financial balanced scorecard measures</b>										
Aug 2002 - Jul 2003	14676 (2405)***	3500 (767)***	2385 (911)***	475 (152)***	251 (162)	740 (103)***	37 (77)	-21 (45)	-65 (13)***	8 (21)
Aug 2003 - Jul 2004	14266 (2263)***	5188 (771)***	2980 (860)***	822 (133)***	580 (171)***	1057 (67)***	130 (72)*	-91 (39)**	-78 (12)***	-0 (36)
Aug 2004 - Jul 2005	13583 (1960)***	3985 (570)***	680 (663)	1182 (113)***	838 (126)***	1186 (61)***	31 (53)	-84 (33)**	-99 (8)***	6 (22)

Note: Robust standard errors in parentheses. 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 brand A branch and the geographically nearest division 2 brand C 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 8a: Matched/Diff-in-Diff results on all variables - branches scoring below median on financial measures in pre-Aug2002 period**

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
<b>3408 observations on 48 branches below median on non-financial balanced scorecard measures</b>										
Aug 2002 - Jul 2003	1665 (3533)	-564 (965)	-418 (1227)	-385 (213)*	-304 (211)	552 (127)***	20 (87)	-25 (57)	-49 (14)***	14 (27)
Aug 2003 - Jul 2004	678 (3301)	-231 (956)	-147 (1027)	-18 (207)	-550 (206)***	828 (124)***	53 (80)	-35 (57)	-77 (14)***	47 (38)
Aug 2004 - Jul 2005	-7080 (3131)**	-1369 (929)	-2269 (1023)**	-151 (195)	105 (177)	940 (96)***	96 (134)	-164 (47)***	-73 (20)***	-8 (26)
<b>2,130 observations on 30 branches above median on non-financial balanced scorecard measures</b>										
Aug 2002 - Jul 2003	19833 (4035)***	4001 (972)***	2605 (1352)*	490 (266)*	183 (296)	730 (164)***	19 (108)	118 (73)	-48 (17)***	4 (23)
Aug 2003 - Jul 2004	20045 (3951)***	5155 (984)***	2626 (1225)**	981 (249)***	296 (213)	1077 (110)***	75 (123)	25 (64)	-47 (19)**	-22 (26)
Aug 2004 - Jul 2005	16018 (3365)***	3238 (708)***	-546 (986)	1065 (195)***	837 (230)***	1308 (90)***	-61 (92)	-69 (53)	-85 (13)***	-14 (14)

Note: Robust standard errors in parentheses. 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 brand A branch and the geographically nearest division 2 brand C 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 8b: Matched/Diff-in-Diff results on all variables - branches scoring above median on financial measures in pre-Aug2002 period**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep var: difference in	Sales	Gross profits	Trading profits	Labour costs	Infrastructure	Transport costs	General and administration	Other	IT	Marketing and advertising
<b>2,130 observations on 30 branches below median on non-financial balanced scorecard measures</b>										
Aug 2002 -	2275	-216	-1428	388	287	853	17	34	-61	23
Jul 2003	(4124)	(1197)	(1446)	(270)	(311)	(168)***	(145)	(80)	(19)***	(29)
Aug 2003 -	7908	283	-2291	1417	316	1196	171	47	-83	29
Jul 2004	(3779)**	(1298)	(1506)	(244)***	(307)	(143)***	(123)	(70)	(19)***	(34)
Aug 2004 -	10617	1148	-3601	2016	1128	1406	231	111	-99	-18
Jul 2005	(3337)***	(982)	(1198)***	(211)***	(482)**	(140)***	(90)**	(62)*	(13)***	(28)
<b>3,408 observations on 48 branches above median on non-financial balanced scorecard measures</b>										
Aug 2002 -	11454	3195	2265	466	291	746	48	-107	-75	11
Jul 2003	(2929)***	(1079)***	(1212)*	(181)**	(189)	(133)***	(105)	(57)*	(17)***	(31)
Aug 2003 -	10631	5227	3227	722	756	1044	166	-165	-97	14
Jul 2004	(2678)***	(1078)***	(1158)***	(149)***	(245)***	(85)***	(89)*	(50)***	(16)***	(56)
Aug 2004 -	11981	4442	1444	1254	834	1110	89	-94	-109	18
Jul 2005	(2371)***	(810)***	(880)	(137)***	(147)***	(82)***	(64)	(42)**	(11)***	(35)

Note: Robust standard errors in parentheses. 3,672 observations on 51 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 brand A branch and the geographically nearest division 2 brand C 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**

	(1) Difference in sales with nearest plumb	(2) Difference in gross profit with nearest plumb	(3) Difference in trading profit with nearest plumb	(4) Difference in labour cost with nearest plumb	(5) Difference in infrastructure with nearest plumb	(6) Difference in transport costs with nearest plumb
<i>[interaction is evaluated at the mean level of service = 6.6 years]</i>						
service x Aug02Jul03	1724 (363)***	735 (107)***	401 (119)***	183 (21)***	80 (31)**	52 (14)***
service x Aug03Jul04	3181 (353)***	1250 (117)***	659 (122)***	230 (25)***	168 (26)***	76 (12)***
service x Aug04Jul05	3857 (420)***	1023 (121)***	459 (130)***	248 (29)***	122 (31)***	58 (14)***
Aug 2002 - Jul 2003	8320.609 (1816.261)***	1514.562 (548.967)***	771.632 (664.487)	192.991 (113.290)*	91.801 (119.922)	703.843 (72.523)***
Aug 2003 - Jul 2004	8939.784 (1682.480)***	2573.438 (552.524)***	998.047 (616.013)	677.274 (103.234)***	184.193 (121.489)	1014.065 (57.878)***
Aug 2004 - Jul 2005	6938.202 (1512.552)***	1838.946 (448.213)***	-1016.917 (516.483)**	934.714 (91.864)***	666.945 (125.289)***	1154.239 (50.343)***

Note: Robust standard errors in parentheses. 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 brand A branch and the geographically nearest division 2 brand C 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 10: Years of service of senior staff**

	(1) Difference in sales with nearest plumb	(2) Difference in gross profit with nearest plumb	(3) Difference in trading profit with nearest plumb	(4) Difference in labour cost with nearest plumb	(5) Difference in infrastructure with nearest plumb	(6) Difference in transport costs with nearest plumb
<i>[interaction is evaluated at the mean level of service = 6.6 years]</i>						
service x Aug02Jul03	798 (433)*	377 (124)***	110 (143)	94 (24)***	105 (36)***	33 (16)**
service x Aug03Jul04	2458 (411)***	724 (129)***	192 (140)	127 (28)***	184 (32)***	49 (14)***
service x Aug04Jul05	3026 (473)***	535 (140)***	48 (157)	111 (30)***	148 (41)***	32 (16)**
<i>[interaction is evaluated at the mean level of service = 11.3 years]</i>						
senior service x Aug02Jul03	589.236 (146.320)***	228.051 (44.486)***	185.144 (50.396)***	56.479 (8.919)***	-16.040 (8.016)**	11.958 (6.797)*
senior service x Aug03Jul04	460.188 (148.061)***	334.536 (52.536)***	297.151 (57.476)***	65.685 (8.965)***	-10.119 (8.848)	17.220 (5.151)***
senior service x Aug04Jul05	528.470 (168.709)***	310.504 (52.602)***	261.701 (57.636)***	87.282 (10.062)***	-16.621 (10.726)	16.782 (5.127)***
Aug 2002 - Jul 2003	8317.042 (1818.743)***	1513.671 (549.294)***	770.852 (664.568)	193.107 (113.358)*	91.837 (119.905)	703.808 (72.506)***
Aug 2003 - Jul 2004	8932.893 (1684.153)***	2569.558 (549.788)***	994.651 (613.455)	676.582 (103.234)***	184.339 (121.502)	1013.869 (57.857)***
Aug 2004 - Jul 2005	6932.842 (1512.138)***	1837.945 (447.028)***	-1017.793 (515.718)**	935.032 (91.249)***	666.987 (125.417)***	1154.207 (50.314)***

Note: Robust standard errors in parentheses. 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 brand A branch and the geographically nearest division 2 brand C 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 A.1**

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_job	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_job	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>

**Table A.2**

Job	Specific job titles included
Branch Manager	Branch Manager, Branch Manager (Designate), Heavyside Manager, Senior Branch Manager
Office	Administration Assistant, Administration Supervisor, Administrator, Assistant Branch Manager, Assistant Depot Manager, Branch Supervisor, Deputy Manager, Estimator, General Clerk, Typist, Office Manager, Pt General Clerk, Secretary, Temp General Clerk
Sales	Contracts/Sales Administrator, Credit Controller, Goods Inwards Assistant, Inside Sales, Pt Sales Clerk, Pt Sales Counter Supervisor, Pt Showroom Sales Asst, Sales Clerk, Sales Counter Assistant, Sales Counter Supervisor, Sales Negotiator, Sales Representative, Sales Supervisor, Showroom Sales Assistant, Showroom Sales Supervisor, Showroom Supervisor, Stock Controller
Driver	Driver, Driver LGV, Glass Cutter/Driver, Warehouse Assistant/Driver
Warehouse/Yard	Depot Manager, Depot Manager Designate, Drainage Supervisor, Foreman, Heavy Supervisor, Hire Assistant, Ironmongery Supervisor, Lightside Supervisor, Logistics Manager, Maintenance Supervisor, Office Supervisor, Operations Manager, Plumbing Supervisor, Product Supervisor, Pt Stores/Warehouse Assistant, Saturday Assistant, Shop/Yard Assistant, Stores/Warehouse Assistant, Temp Yard Assistant, Timber Supervisor, Transport Supervisor, Warehouse Supervisor, Yard Assistant, Yard Assistant/Driver, Yard Manager, Yard Supervisor, Yard/Warehouse Assistant, Yard/Warehouse Supervisor
Other	Fixer, Machine Operator, Mill Operative, Pt Cleaner, Timber Machinist, Trainee B2

Figure 1: Firm Structure

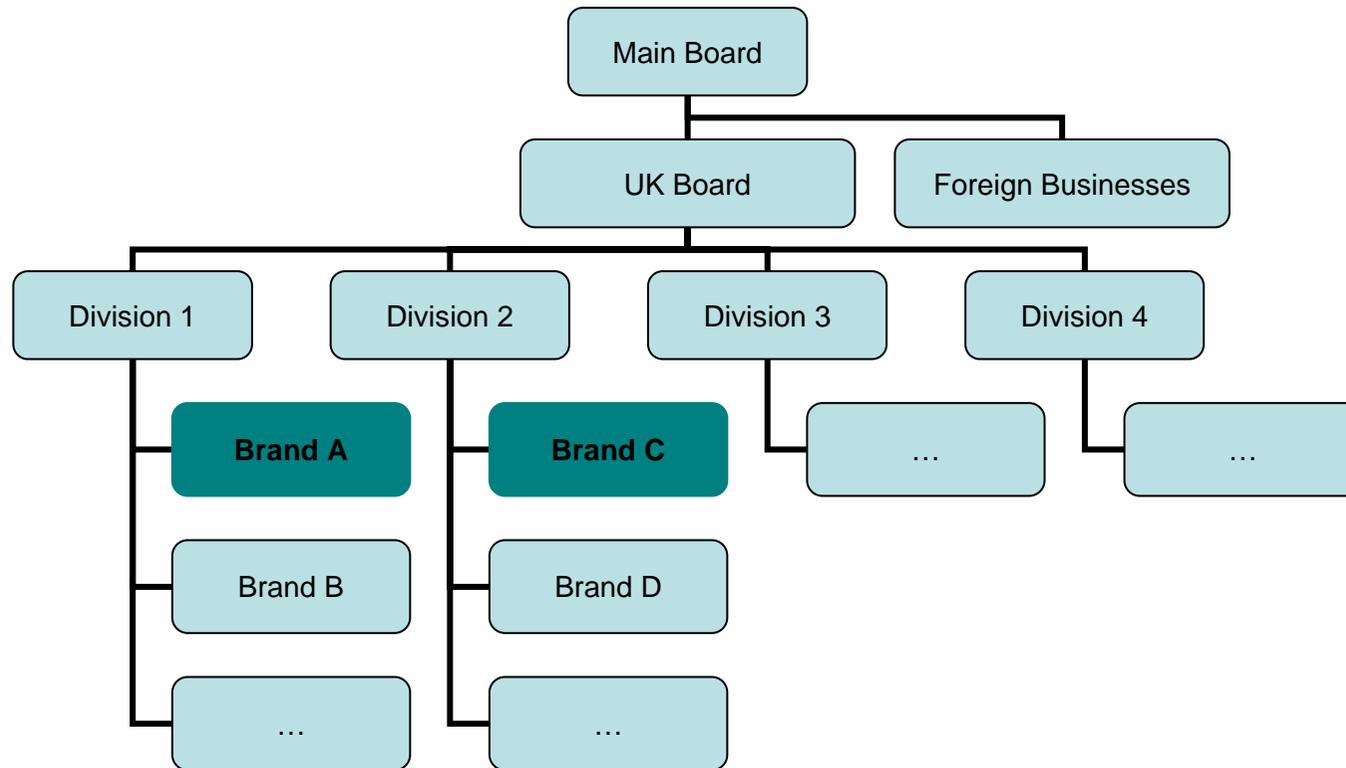


Figure 2: The Scorecard measures

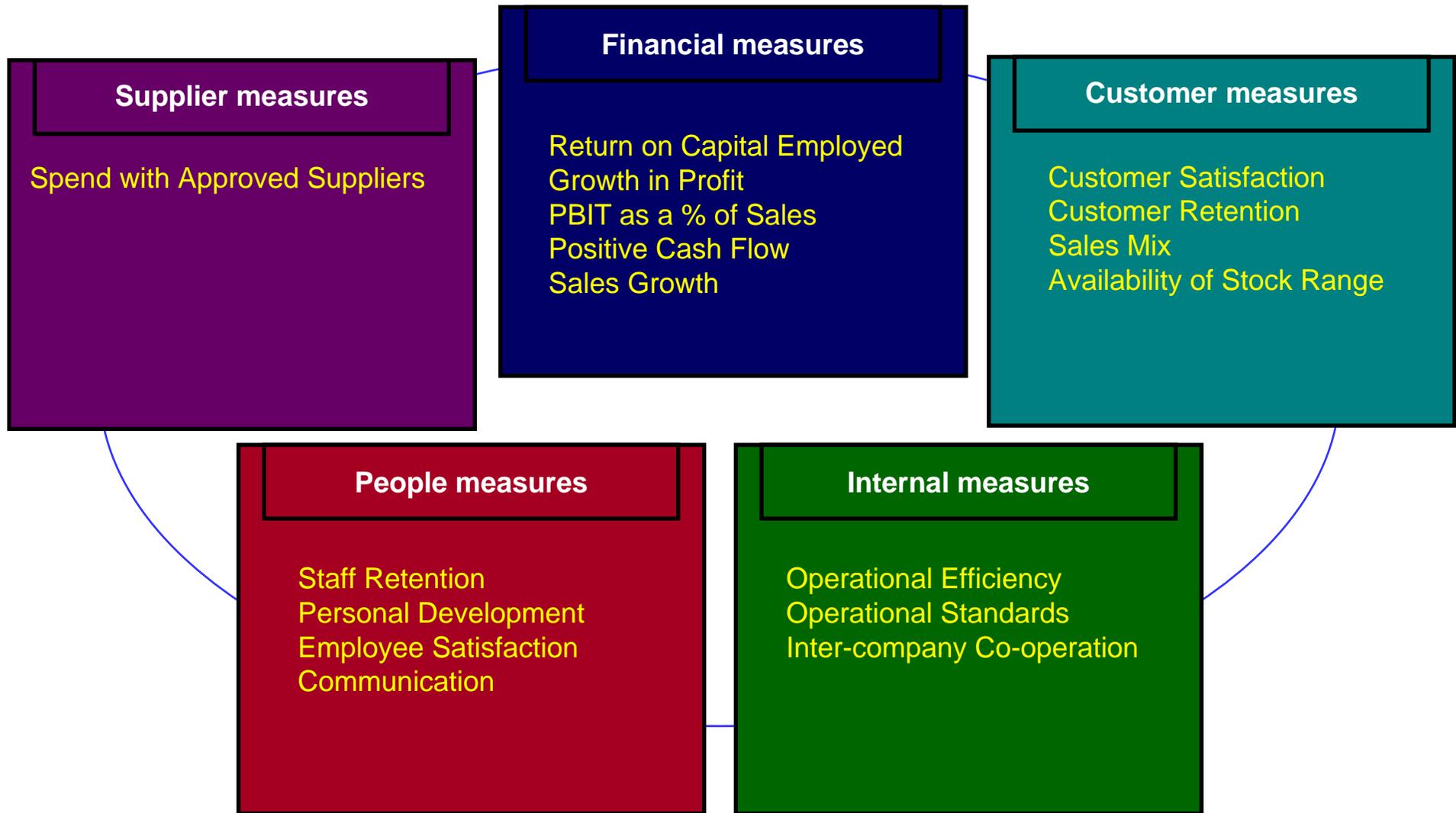
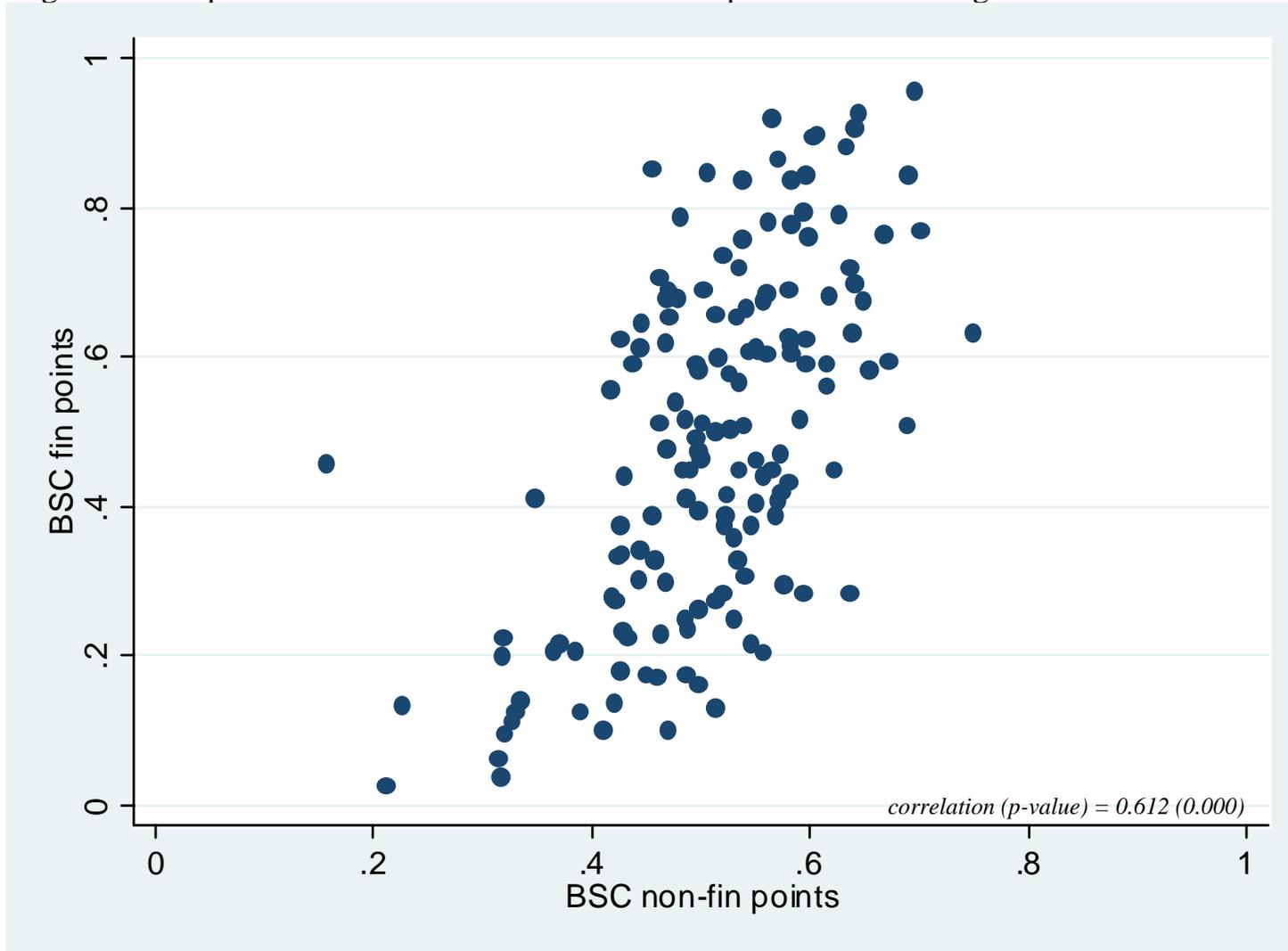
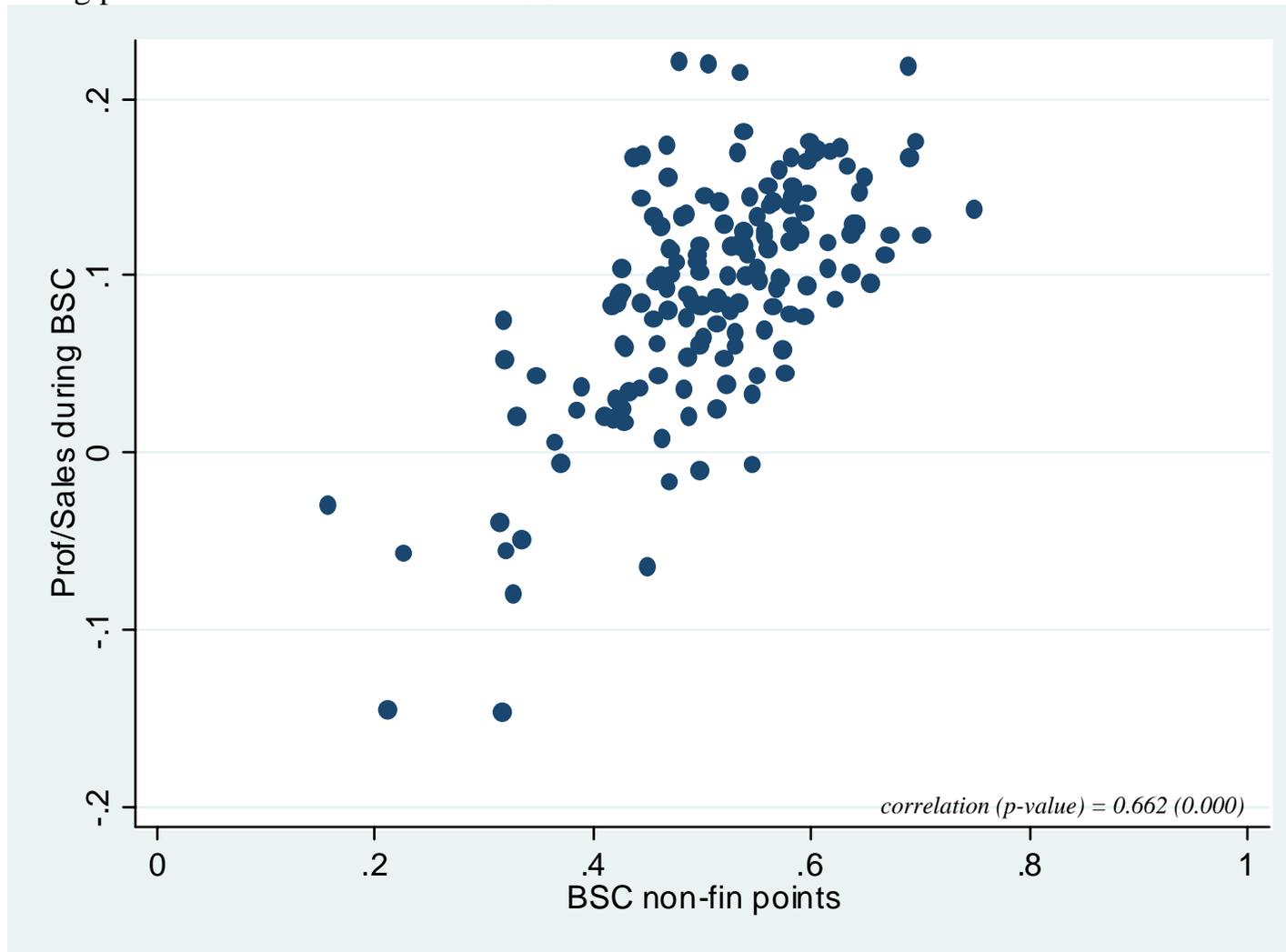


Figure 3: Comparison of non-financial with financial performance during Balanced Scorecard



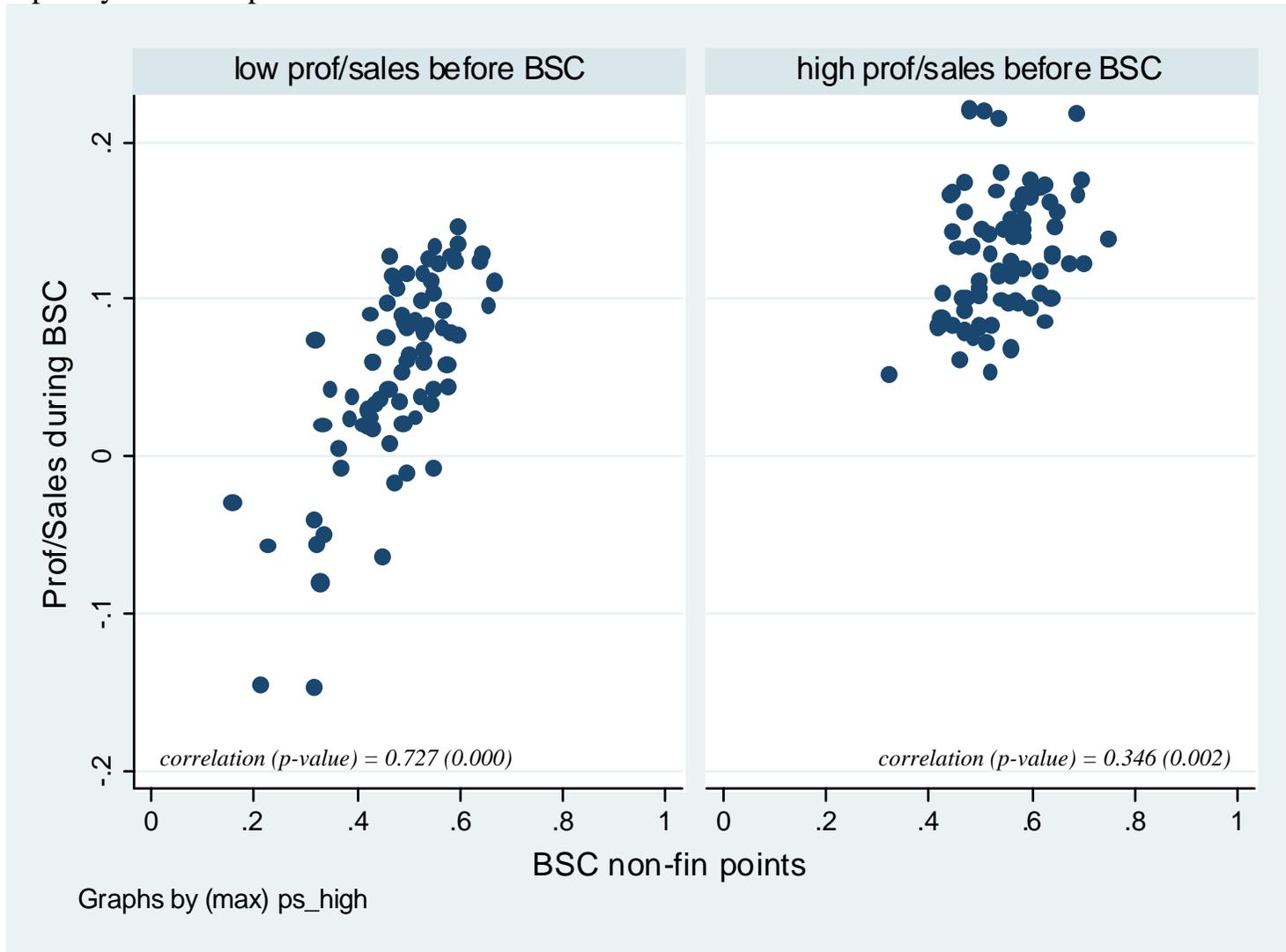
Notes: 156 observations (one for each Division 1 Brand A branch); x-axis is the share that each branch earned of the total points they could have earned on the non-financial BSC measures (customer, internal, people and supplier) over the period August 2002 - July 2004; y-axis is the share that each branch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

Figure 3b: Comparison of non-financial with financial performance during Balanced Scorecard, using profits over sales rather than BSC measures



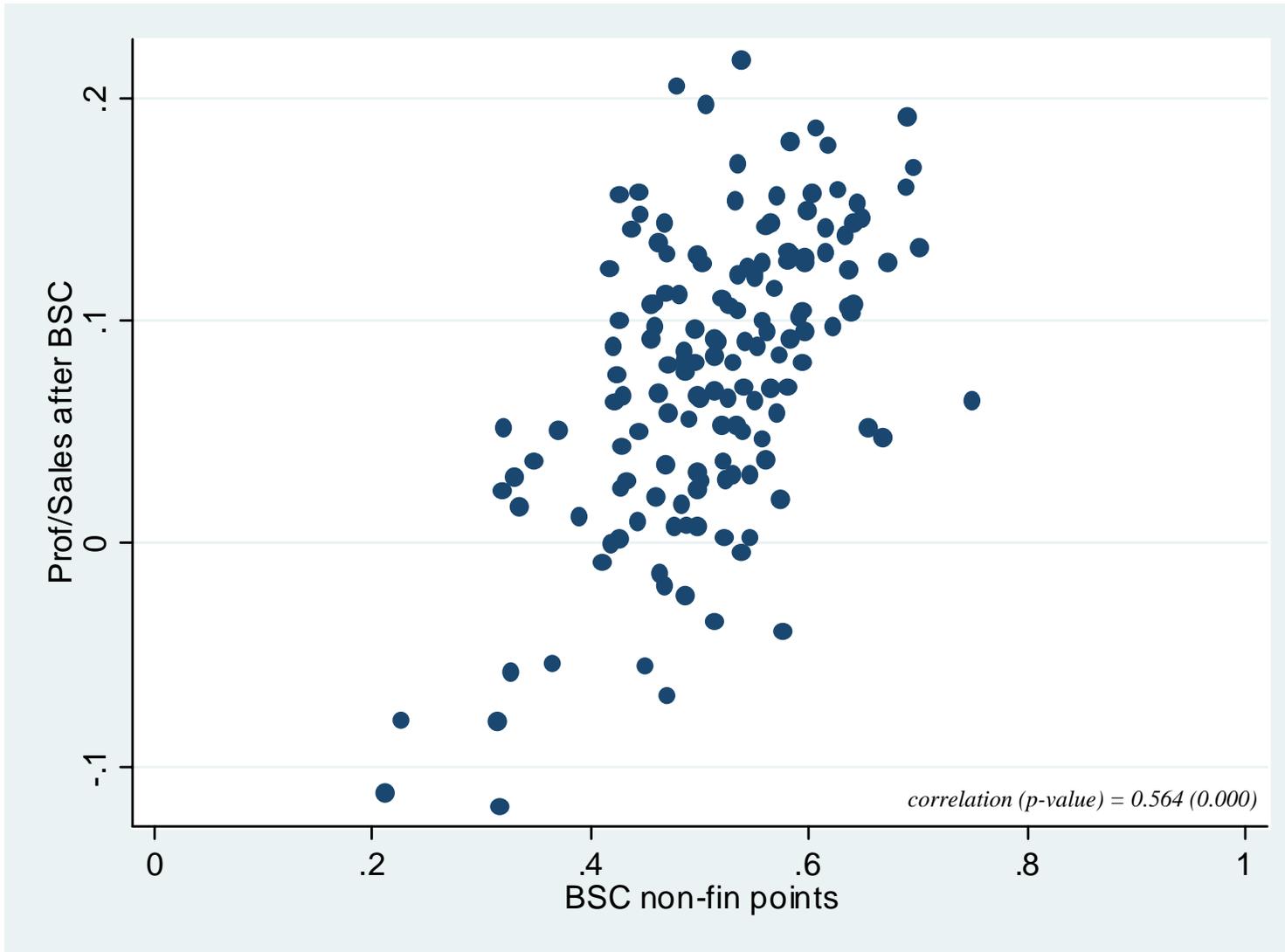
Notes: 156 observations (one for each Division 1 Brand A branch); x-axis is the share that each branch earned of the total points they could have earned on the non-financial BSC measures (customer, internal, people and supplier) over the period August 2002 - July 2004; y-axis is the average value of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

Figure 4: Comparison of non-financial with financial performance during Balanced Scorecard, split by financial performance before BSC



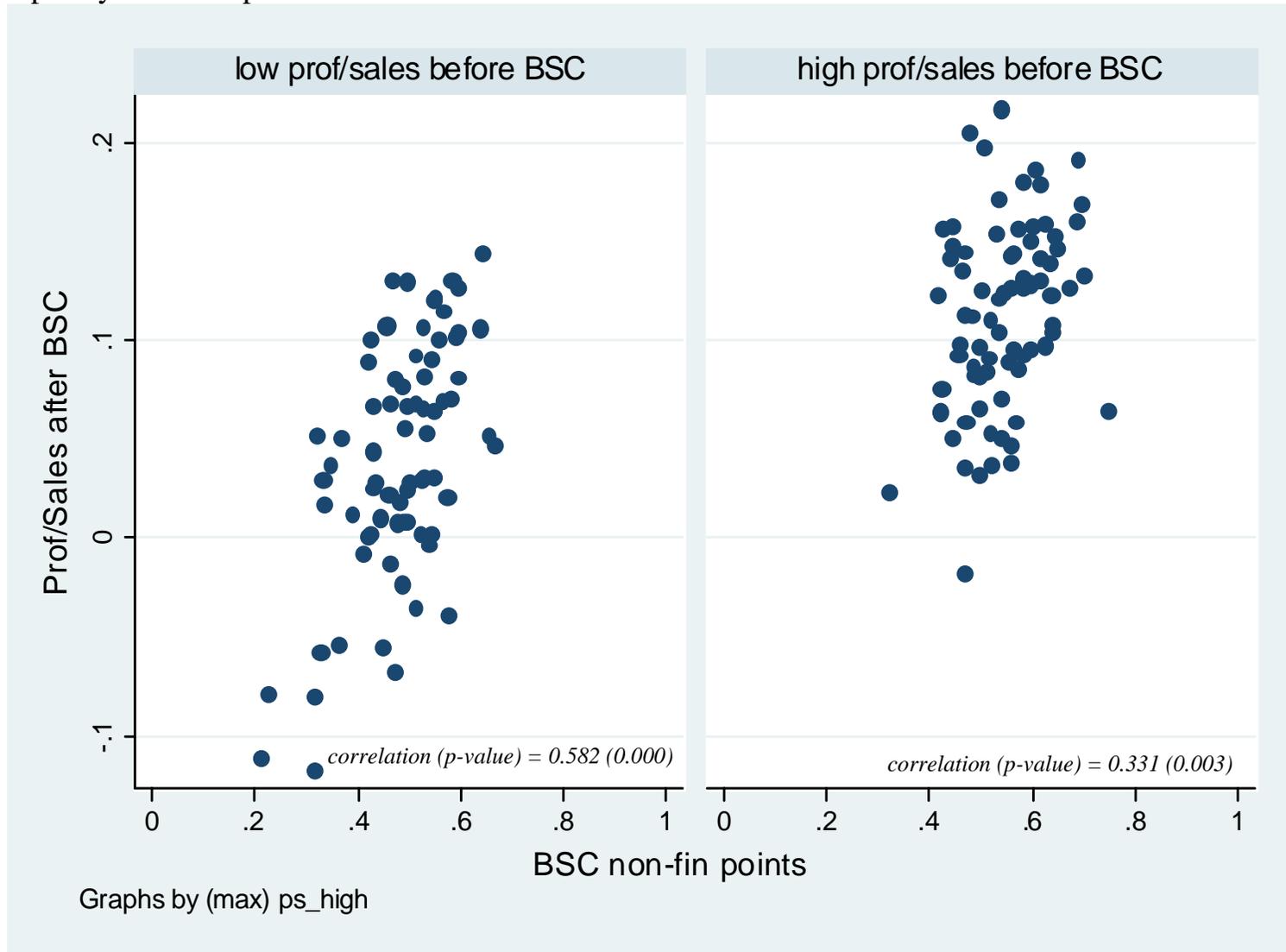
Notes: 156 observations (one for each Division 1 Brand A branch); x-axis is the share that each branch earned of the total points they could have earned on the non-financial BSC measures (customer, internal, people and supplier) over the period August 2002 - July 2004; y-axis is the average value of 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 5: Comparison of non-financial with financial performance after Balanced Scorecard



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 2004 - July 2005.

Figure 6: Comparison of non-financial with financial performance after 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 2004 - July 2005; left-hand panel are those branches that had below median ratio of profit to sales over the period August 1999 - July 2002.