

VU Research Portal

Non-financial performance measures: an empirical analysis of a change in a firm's performance measurement system

Wiersma, E.

2003

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Wiersma, E. (2003). *Non-financial performance measures: an empirical analysis of a change in a firm's performance measurement system*. Thela Thesis (Tinbergen Instituut).

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

NON-FINANCIAL PERFORMANCE MEASURES: AN EMPIRICAL ANALYSIS
OF A CHANGE IN A FIRM'S PERFORMANCE MEASUREMENT SYSTEM

ISBN 90 5170 669 3

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 298 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

VRIJE UNIVERSITEIT

NON-FINANCIAL PERFORMANCE MEASURES: AN EMPIRICAL ANALYSIS
OF A CHANGE IN A FIRM'S PERFORMANCE MEASUREMENT SYSTEM

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. T. Sminia,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Economische Wetenschappen en Bedrijfskunde
op woensdag 16 januari 2003 om 13.45
in het auditorium van de universiteit,
De Boelelaan 1105

door

Eelke Wiersma

geboren te Dokkum

Promotor: prof.dr. T.L.C.M. Groot

Preface

Chapter 1

Introduction	1
1.1 Introduction	1
1.2 Non-financial performance measures	4
1.3 Management control	6
1.4 Managerial performance	9
1.5 Contributions of the study	10
1.6 Summary and overview of remainder dissertation	12

Chapter 2

Literature review	13
2.1 Introduction	13
2.2 Overview empirical research	14
2.3 Relationships between financial and non-financial performance measures	15
2.3.1 Empirical evidence of relationships between financial and non-financial performance measures	15
2.3.2 Characteristics of the relationship between financial and non-financial measures	17
2.4 Antecedents of non-financial performance measures use	18
2.4.1 Introduction	18
2.4.2 Operational strategy	18
2.4.3 Organizational strategy	20
2.4.4 Other antecedents	21
2.4.5 Summary	23
2.5 Performance effects of non-financial performance measure use	24
2.6 Discussion and future research	27
2.6.1 Relationships between financial and non-financial performance measures	27
2.6.2 Use and usefulness of non-financial performance measures	30
2.7 Summary and conclusions	33

Chapter 3

Research site	35
3.1 Introduction	35
3.2 Organization	35
3.3 Change in the PMS	37
3.4 Goals of the change in the PMS	37
3.5 Performance measurement cycle	38
3.6 Controllability of measures by area manager	39

Chapter 4

Data description	41
4.1 Introduction	41
4.2 Variable definition	42
4.3 Pooled time-series data	44
4.3.1 Introduction	44
4.3.2 Properties of pooled time-series data	44
4.3.3 Fixed effects model	46
4.4 Description of the data	47

Chapter 5	
Relative and incremental information content	49
5.1 Introduction	49
5.2 Framework	50
5.2.1 Relative information content	50
5.2.2 Incremental information content	51
5.3 Specification	53
5.3.1 Relative information content	53
5.3.2 Incremental information content	55
5.3.3 Statistical considerations	57
5.4 Empirical results	58
5.4.1 Introduction	58
5.4.2 Relative information content	58
5.4.3 Lag search	60
5.4.4 Incremental information content	61
5.4.5 Managerial significance	66
5.4.6 Non-linearity in incremental information content	67
5.5 Discussion and summary	69
Chapter 6	
Performance effects of including non-financial measures in the contracts of managers	71
6.1 Introduction	71
6.2 Framework	72
6.2.1 Introduction	72
6.2.2 Agency theory	72
6.2.3 The principal-agent problem and management accounting procedures	73
6.2.4 Value of non-financial performance measures in the multi-task principal-agent framework	74
6.2.5 Limitations of the principal-agent model	75
6.2.6 Effect of change in PMS on financial and non-financial measures	75
6.3 Research design	77
6.3.1 Introduction	77
6.3.2 Quasi-experimental design	77
6.3.3 Interrupted time-series design	78
6.3.4 Threats to internal validity	79
6.4 Research methods	80
6.4.1 Introduction	80
6.4.2 Time-series regression versus time-series analysis	80
6.4.3 Impact assessment	81
6.4.4 Impact assessment in accounting research	83
6.5 Specification	84
6.5.1 Statistical consideration	86
6.6 Empirical results	87
6.6.1 Introduction	87
6.6.2 Impact of change in PMS with yearly data	87
6.6.3 Impact of change in PMS on non-financial performance	89
6.6.4 Impact of change in PMS on financial performance	90
6.6.5 Managerial significance	91
6.7 Summary and conclusion	91

Chapter 7	
Summary and conclusion	93
7.1 Introduction	93
7.2 Summary	93
7.3 Conclusions and implications	95
7.3.1 Information content of performance measures	95
7.3.2 Evaluation of the change in the PMS	96
7.4 Contributions of the dissertation	97
7.5 Limitations of the study	98
7.6 Further research	99
Appendix: Graphs of data	101
References	105
Summary in Dutch	115

Preface

Although most prefaces of dissertations are larded with comments about the amount of work it took to finish it, I would like to stress that most of the time I felt privileged to be able to do so. Trips to conferences, analyzing data and reading illuminating papers more than compensates the sometimes laborious work in the office and not so encouraging comments of reviewers.

I thank my supervisor Tom Groot for his enthusiastic support and suggestions during all phases of my Ph. D. project. I also thank my dissertation comity Frank Hartmann, Siem Jan Koopman, Fred Lindahl, David Otley, and Marc Wouters for their helpful remarks on the manuscript.

The accounting department of the Vrije Universiteit Amsterdam is a very friendly environment to work in, for this I want to thank all my colleagues. Specially, I want to thank Erik, Henri, Ivo, Maarten, and Martijn for all the discussions we had about accounting, statistics, and research methodology that facilitated me to execute the dissertation and made me enjoying my job.

Six months of my Ph. D. period I stayed in Dallas, which was a terrific and insightful time. I want to thank Rajiv Banker and his research group for giving me this opportunity.

This dissertation would never had exist when the company that supplied the data would not have cooperated. I want to thank everyone from all layers in the company for their cooperation and time spent.

Finally, I want to thank my family for all their unspoken support. Especially, I want to thank Janine for sharing her life with me in such a pleasant way.

Chapter 1 Introduction

1.1 Introduction

The aim of this dissertation is to assess the usefulness of using non-financial performance measures in the performance measurement and evaluation system of middle managers. It reports two empirical studies in a large Dutch logistics organization that had a substantial change in its performance measurement system (PMS). More specific, a change in the PMS took place, where before the change managers were evaluated mainly on financial measures, whereas after the change additional non-financial measures were included in the performance evaluation system. In the first study, non-financial and financial performance measures are compared to assess their ability to predict future financial performance. The second study assesses the impact of the change in the PMS on managerial performance through a quasi-experiment.

Since the introduction of the Balanced Scorecard concept (Kaplan and Norton, 1992), the subject of non-financial performance measurement has received a lot of attention from both practitioners and academics. The Balanced Scorecard (BSC) and other comparable performance measurement systems were developed as a reaction to the general dissatisfaction of using only financial measures for performance measurement and evaluation purposes. Financial measures were argued, among others, to be too aggregated, to be too focussed on the short-term, and to reflect only effects of actions and not about causes. All advantages given of non-financial performance measures can be summarized as follows. Managerial efforts lead to future outcomes, but these outcomes are not all reflected in current financial measures. Therefore, the non-financial measures should help managers to refocus their attention on tasks that drive these future outcomes (Hemmer, 1996).

Banker et al. (2001) show from a sample of 165 financial executives from the US, that 35% used the BSC. In Europe, a survey of 236 Scandinavian business unit managers indicated that 27% used a BSC (Kald and Nilsson, 2000). In contrast, several sources show that managers have many problems implementing Balanced Scorecard systems (for example, Ittner and Larcker, 1998b). This shows the importance of the topic among practitioners.

Ittner and Larcker (1998b) derive a number of researchable issues with respect to non-financial performance measurement. First, whether the use of non-financial performance measures, or concepts such as the BSC that incorporate non-financial measures, has positive benefits for organizations. Second, when this is not the case for all organizations, identify variables that moderate the relationship between using more non-financial performance measures and increased performance is an important issue. Finally, other addressable issues are what does the concept of balance mean in the BSC, how a number of different performance measures should be used in the evaluation of managers, and whether the same measures can be used for both improved decision making and control.

In this dissertation two general research questions are addressed. First,

“Do non-financial performance measures have relative or incremental information content, or both, beyond financial performance measures in predicting future financial performance?”

And second,

“Do managers perform better when non-financial measures are added to their performance evaluation system?”

The first research question examines a claim in the popular management accounting literature. Kaplan and Norton (1992) argue that current non-financial measures are better indicators for future financial performance than current financial measures. In addition, Banker et al. (2000b, p. 65) argue that:

“current non-financial measures are better predictors of long-term financial performance than current financial measures” (italics added).

Finally, Nagar and Rajan (2001, p. 496) state in the context of quality cost systems, that:

“critics of traditional quality cost systems propose supplementing financial measures with non-financial quality measures, arguing that non-financial measures provide a better indication of quality related customer goodwill losses” (italics added).

These claims are based on the conjectures that non-financial measures, compared to financial measures, for example, pay more attention to causes rather than effects (Singleton-Green, 1993), and induce a long-term orientation in managers (Kaplan and Norton, 1992). These claims suggest that non-financial measures have a higher relative information content than financial measures for predicting future financial performance, i.e., that non-financial measures contain more information than financial measures (Biddle et al., 1995).

However, empirical accounting research assesses the usefulness of non-financial measures through the incremental information content of non-financial measures beyond financial measures for predicting future financial performance. For example, Ittner and Larcker (1998b, p. 226) argue that¹:

¹ However, Ittner and Larcker (1998b) also state that predictive ability of non-financial measures does not necessarily mean usefulness for contracting, since the weight of a measure is also influenced by the noise of the measure.

“from an accounting standpoint, a crucial test is whether a broad set of non-financial measures [...] possess *incremental* ability to predict future financial performance, after controlling for the predictability of past financial performance” (italics added).

This remark hints that non-financial performance measures should have “incremental information content”, that is they provide information beyond the lagged financial measures (Biddle et al., 1995). In addition, Ittner and Larcker (1998a), Banker et al., (2000b), and Nagar and Rajan (2001) test the incremental information content of non-financial measures beyond financial measures for future financial performance. The accounting literature therefore is ambiguous whether non-financial measures have relative or incremental information content, or both, to predict future financial performance beyond lagged financial measures. Therefore, in this dissertation I test both the relative and incremental information content of non-financial measures to explore this contention.

In addition, a critical assumption in the BSC concept is that non-financial measures are indicators for financial performance. The first research question also addresses this assumption. Non-financial measures do not need to be an indicator for current financial performance, but only for future financial performance. Prior studies that addressed this question used either an arbitrary lag (Anderson et al., 1994; Ittner and Larcker, 1998a; Nagar and Rajan, 2001) or had the same lag for all non-financial measures and all financial measures (Banker et al., 2000b).² Therefore, in this study a lag search is executed to explore the lag length between non-financial measures and future financial measures, in which different non-financial measures can have different lags for different financial measures.

Three years of monthly performance data on two non-financial and two financial measures is available for 27 areas of one company to examine the first research question. These measures are the measures from the contracts of middle managers. The non-financial measures are on-time delivery and worker satisfaction, and the financial measures are costs and revenues.

The second research question is an evaluation of the change in the PMS. Foster and Young (1997) argue that much of the management accounting literature is larded with proposals of new accounting systems, but that any systematic evaluation of such new systems is scarce.³ This is in sharp contrast with evaluations of new products or programs in other disciplines, such as medicine, and education, in which each new initiative is tested before use. The second study in this dissertation is an empirical account of such an evaluation in one organization. Although the new PMS was not a BSC, the goals of the change were similar.

Available empirical research for the second research question provides mixed evidence for a number of reasons. First, the outcome measures used to indicate the success of emphasizing non-financial measures relatively more are short-term outcome measures that do not incorporate the full performance effects (Perera et al., 1997). Second, most studies use a cross-sectional methodology and therefore do not capture long-term performance effects of emphasizing non-financial performance measures relatively more (Perera et al., 1997). Therefore, the second research question in this study is explored through a quasi-experiment in a large Dutch service

² It is not clear from Banker et al. (2000b) whether this same lag between the two non-financial measures and three financial measures was a restriction in the analysis or a coincidence. Personal communication with the first author learned that it was a coincidence.

³ Notable exceptions are Banker et al. (1996a, 1996b, 2000b) and Emsley (2000).

organization where a change in the performance measurement system (PMS) has taken place. Before the change in the PMS managers were mainly evaluated on financial measures, whereas after the change additional non-financial measures were emphasized relatively more.

To answer the second research question I use the same data as for the first research question. Of these three years, one year is from before, and two years are from after the change in the PMS.

Both research questions are not independent of each other. The first question might be considered a validity check whether the added non-financial measure were value drivers in the context of the organization. For example, the informativeness hypothesis implies that all indicators that have incremental information content about the effort of a manager should have a non-zero weight in the contract of managers (Ittner and Larcker, 1997). When this is not the case no performance effects can be expected of the change in the PMS.

The remainder of this chapter is organized as follow. First, the subject of non-financial performance measurement is outlined and defined. Second, the subject of non-financial performance measurement is situated in the management control literature. Third, the evaluation question uses managerial performance as outcome measure. This is defined in section 1.4. Fourth, the contributions of the study to the literature are enumerated. Finally, an overview of the dissertation is given.

1.2 Non-financial performance measures

Labels for performance measures that are alternatives to financial or accounting measures abound. In this section I discuss a number of alternative labels used in the literature that are partly overlapping with the financial and non-financial dimension. In addition, I give advantages and disadvantages of non-financial performance measures.

The American Accounting Association defines financial information as:

“a quantitative measure, expressed in the monetary metric, resulting from the measurement of past, present and future economic events, or that has a financial character (American Accounting Association, 1975)”.

Morisette (1998, p. 4) derives from this definition a definition for non-financial measures⁴:

“any quantitative measure, 1) expressed in a metric other than a monetary unit, or 2) that results from mathematical manipulations or ratios of pieces of information expressed in metrics other than monetary units”.

These definitions suggest that the basic content of the difference between financial and non-financial measures is the unit of the measure, i.e., financial versus other units. However, the accounting literature encloses numerous other labels that are partly overlapping with the financial versus non-financial distinction. First, the distinction between accounting and non-accounting information is central in the “reliance on accounting performance measure” construct. However, it is not clear what the exact contents of the accounting and non-accounting categories are (Hartmann, 2000). In

⁴ Horngren et al. (1994) make a further distinction between internal and external financial and non-financial measures.

addition, Abernethy and Lillis (1995, p. 242) refer to “accounting and other efficiency based measures”, but the items that measure this construct are a number of both financial and non-financial measures. Further, in the Balanced Scorecard the notion of leading versus lagging measures is central, in which the lagging measures are most often considered to be the financial measures and the leading measures are considered to be the non-financial measures.⁵ Finally, Govindarajan and Gupta (1985) use the construct long-run versus short-run criteria in the bonus system, but again the items in the construct used are financial measures for the short-run criteria and non-financial measures for the long-run criteria. The existence of these different labels raises the question whether the distinction between giving measures in financial or non-financial units is substantial, and what the basic argument is for this difference.^{6,7}

To make the issue even more complex, some empirical evidence suggests that financial and non-financial measures are not substitutes, but that non-financial measures are used additive to financial measures (Govindarajan and Gupta, 1985). This implies that financial and non-financial measures are not the endpoints of a continuum, and therefore a scale that uses these endpoints cannot measure the dimension properly.

Given the diverse labels that exist in the literature it is difficult to give a clear-cut definition of financial and non-financial performance measure that captures the real difference. Therefore, acknowledging the problematic nature of the definitions, I use the definitions of financial and non-financial measures given by the American Accounting Association (1975) and Morisette (1998) given above. This implies that I only consider quantitative measures, and therefore qualitative non-financial measures are not part of the definition.

The literature provides several claims in front of emphasizing non-financial performance measures relatively more. A non exclusive enumeration is that non-financial performance measures are argued to be more actionable, be more directly traceable to the strategies of the firm, work well with high-technology manufacturing systems like Automatic Manufacturing Technologies and JIT, facilitate continuous improvements (Fisher, 1992), focus attention on causes rather than effects (Singleton-Green 1993)⁸, focus attention on the long-run perspective (Kaplan and Norton, 1992), and give less room for dysfunctional behavior (Fisher, 1992). Empirical evidence for these advantages is scant. Van der Stede et al. (2001) find that non-financial measures encourage risk taking and innovation, have less noise, and are less susceptible to short-termism and gamesmanship. Disadvantages of non-financial performance measures are stated less often. This list might include that they are not always directly

⁵ However, in Nagar and Rajan (2001) financial quality costs, i.e., internal failure costs and external failure costs, are also indicators for future revenues. This example shows that the distinction between leading and lagging is not the same as the distinction between financial and non-financial measures.

⁶ Although deriving proper labels for a construct is a theoretical question, empirically the difference between the different labels and the discriminant validity of the constructs could be assessed through use of factor analysis (Hair et al., 1998).

⁷ The coexistence of all these different labels illustrates the risks of using “practice defined variables” (Luft and Shields, forthcoming). For example, although accounting and non-accounting measures are labels that are used for many years, its underlying theoretical properties are still not well understood.

⁸ For example, Singleton-Green (1993, p. 52) argued that “profit measures show the effects of non-financial activities and achievements; they do not pin down precisely what it is in your business that you are getting right or wrong”. In addition, from an experiment Luft and Shields (2000) find that managers are better able to identify the impact of quality problems on future profits when the quality problems are reported in non-financial terms than in financial terms.

linked with financial measures⁹, are considered to be 'softer' measures, have more measurement error¹⁰, and because different non-financial performance measures are stated in different units, there is not one integrative denominator (Singleton-Green, 1993). Further limitations are that they are not audited, and easier to manipulate (Ittner et al., 1997).

Ittner and Larcker (1998b) give a classification of three reasons why non-financial performance measures are used more often. First, non-financial performance measures try to mitigate limitations in traditional accounting based measures. Traditional accounting measures are, for example, backward looking, emphasize effects instead of causes of problems, are too aggregated, neglect soft or difficult to measure assets, lead to short-termism. Second, due to increased competitive pressure companies need to have more information about the relevant drivers of their performance. Third, operational strategies such as TQM, JIT and flexible production automation lead to a demand for non-financial measures to implement, and trace the effects of these strategies.

The discussion above lists advantages and disadvantages of using non-financial measures. However integrating financial and non-financial measures, for example in a BSC, is argued to give additional advantages. Such integrated systems are assumed to improve the communication of the strategic plan¹¹, help implement the strategic plan, and facilitate strategic learning (Kaplan and Norton, 1996). Van der Stede (2001) gives some empirical evidence for these claims. He finds that integrating financial and non-financial measures adequately captures lead-lag relationships, and reinforces the strategic plan.

1.3 Management control

This section embeds the use of non-financial performance measures in the management control literature. I discuss the concept of (management) control, and the role of non-financial performance measures within management control.

The concept of control is defined in an infinite number of ways in the literature. This study follows Merchant's (1998, p. 2) perspective of control. Therefore, the definition of control used is:

“controls [...] are devices managers use to ensure that the behaviors and decisions of people in the organization are consistent with the organization's objectives and strategies”.

To be able to discuss some critical elements in this definition, I first describe the cybernetic control concept.

Cybernetic control consists of the elements 1) set goals, 2) measure performance, 3) compare performance with the goals, and 4) provide feedback to correct a variance between the performance and the goal. This model assumes that goals are clear, performance can be measured, a predictive model is available, and a set of corrective

⁹ For example, Brancato (1995) reports from a case study that none of the interviewees could quantify the link between non-financial and financial measures.

¹⁰ For example, Strivers et al. (1998) find from a sample of 253 CEO's that 75.9% think employee involvement is very important but only 35.9% measure this dimension because they find it is difficult to measure.

¹¹ Malina and Selto (2002) find that this is the case when the BSC measures and its implementation are credible, the BSC fits in the culture of the company, and BSC leads to increasing and sharing knowledge.

actions is available when goals and actual performance deviate (Otley and Berry, 1980). Although a helpful metaphor for understanding the concept of control, the concept lacks at least three important aspects of controls in organizations.¹² First, organizations act in their environment, and this leads to uncontrollable elements in the goals set. This implies that the predictive model is incomplete and therefore feedback actions do not necessarily lead to a system that is in control. Second, cybernetic controls lack the impact of human behavior, that is individuals have their own goals that may conflict with organizational goals, and individuals have limitations that may inhibit them to execute their tasks perfectly. Again this leads to the result that corrective actions taken do not necessarily lead to a system that will be in control again. Finally, unlike the cybernetic control concept organizations have a number of conflicting, or even vague, goals that are not always compatible with each other. Thus, not only the set of corrective actions available to keep the organization in control is incomplete, but also the goals themselves may be unclear or incorrect.

Merchant's definition also assumes that organizations have objectives and strategies for themselves. The most likely operationalization is that these organizational objectives and strategies are the objectives and strategies of the dominant coalition of the organization (see also section 1.4). In addition, to be able to influence the "behavior and decisions of people" it should be recognized that these people have goals for themselves that may conflict with the organizational goals or that the goals are not clear for them. Finally, Merchant defines the efficacy of controls as the increased probability that the organizational goals are achieved or exceeded (Merchant, 1998). Again this assumes that a clear picture of organizational goals exists.

The performance measurement system from the company under study (see chapter 3) has the same flaws as the cybernetic control concept. First, it provides measures derived from the goals of the company that are partly influenced by the environment of the organization. Second, the managers in the organization might have goals that differ from the organizational goals. Finally, the different performance measures derived from the organizational goals might be conflicting with each other.

Generally, control systems can have two functions, i.e., strategic control and management control (Merchant, 1998). Strategic control questions the validity of the strategies of the organization in a changing environment. In contrast, management control deals with the issue "are the employees behaving appropriately". In this study the focus is mainly on management control, although the change in the PMS of the organization (see chapter 3) might be a consequence of the strategic control process.

The need for management controls stems from control problems. Merchant (1998) classifies these control problems in 1) lack of direction, 2) personal limitations, and 3) motivational problems. Lack of direction means that individuals do not know what the organization expects from them, for example, because the organizational goals are unclear. Personal limitations are the lack of ability of the individual to execute the job due to, for example, inappropriate training or unavailability of proper information. Motivational problems are the most widely stated control problems and are the result of a number of different causes. For example, managers can have different time-horizons compared with the organization, managers are effort and risk averse, and managers spend too much money on, for example, perquisites.¹³

¹² Hartmann (1997) argued that it is not the model itself that is inappropriate, but that the assumptions underlying the model do not hold.

¹³ Goal incongruence, a central concept in multi-task principal-agent models, refers to both lack of direction and motivational problems (Merchant, 1998).

Merchant (1998) classifies control devices in results controls, action controls, and personal and cultural controls. Non-financial performance measurement can best be described as a control device where the performance of managers is measured, evaluated and rewarded on a number of both financial and non-financial measures that are important for the organization. This description includes that relevant performance dimensions are defined, measured, set targets for, and is input for the rewards of managers. Therefore, non-financial performance measurement is a results control device.

Merchant (1998) argued that results controls are able to directly mitigate the control problems lack of direction and motivational problems. However, non-financial performance measures can explicitly help managers learn lead-lag relationships between different dimensions of performance¹⁴ and hence help them to do a better job. This indicates that non-financial performance measures might also lessen the personal limitation control problem. The role of non-financial performance measures in mitigating control problems is summarized in table 1.1.

Table 1.1: Control problems and non-financial measures

Control problems	Role of non-financial measures to address control problems
Personal limitations	Learning lead-lag relations between financial and non-financial measures
Lack of direction	Adding non-financial measures to the contract informs managers what is expected of them
Motivational problems	The non-financial measures reallocate the focus of managers to the non-financial measures by evaluating the managers on these measures

Perera et al. (1997, p. 561) make a distinction between the instrumental versus motivational premise of efficacy of performance measurement¹⁵. The instrumental premise is described as:

“for manager actions and decisions to be effective in achieving [performance dimensions] such actions and decisions need to be informed by relevant and specific feedback on those dimensions”.

The motivational premise is described as:

“managers have an incentive to concentrate on, and will seek to maximize performance against, those activities on which performance is measured”.

These two premises are not only difficult to separate in empirical research but might be potentially conflicting. For the instrumental premise managers need to represent information in an honest way to their evaluators. However, when the same measures

¹⁴ There is some evidence that, for example, CEO's have great difficulties to link improvements in non-financial measures to financial measures. Ittner and Larcker (1998b) surveyed 27 CEO's from large US companies and found that only approximately 25% claimed that they could relate quality and customer satisfaction dimensions of performance to accounting returns.

¹⁵ Other labels in the literature are decision making versus decision control (Zimmerman, 1997), and decision making versus stewardship (Gjesdal, 1981).

are also used for motivational purposes, individuals have an incentive to misrepresent the information.¹⁶ The performance measures in the contracts of the managers in the company also have this tension. The managers use the measures to make tradeoffs between different performance measures but are also evaluated on these same measures.

The change in the PMS considered in the organization (see chapter 3) is installed to align the goals of the managers with the organizational goals, thus mitigating motivational problems. Second, it focuses managers on certain performance dimensions, thus mitigating the problem of lack of direction. Finally, the PMS gives the managers an indication of trade-offs between different performance dimensions.

1.4 Managerial performance

The second research question evaluates whether managers perform better when they are evaluated and rewarded relatively more on non-financial measures. Before this question can be answered, better managerial performance has to be defined.

Ittner and Larcker (2001) discuss the problem of endogeneity when performance is used as a dependent variable. When all organizations optimize their accounting systems no performance effect can be expected of a change in the PMS. However, this assumes that all organizations act rational, and that the changes take effect immediately, which seems highly unlikely. As Milgrom and Roberts (1992) state, it seems more likely that people learn how to make good decisions by experimentation and imitation. The change in the PMS in the organization under study (see chapter 3), was a reaction to the perceived need to react to shocks in the environment. The organization can therefore be seen as temporarily out of the equilibrium. The change in the PMS is a reaction to this. Throughout the change, the organization and its managers tried to adapt the organization to the changing needs of the environment. Thus, with the change in the PMS examined the organization strives for optimality. Consequently, this is tested in the empirical analysis (Banker et al., 1996a; Banker et al., 2000b).

A number of different performance models exist. Meyer and Gupta (1994) enumerate the maximizing model, the political model, the constituency model, and the business model. These models are summarized in table 1.2.

¹⁶ For example, in early publications of Kaplan and Norton about the BSC, they argued that the BSC should not be used for rewarding purposes, since this would give managers an incentive to misrepresent the information that is necessary to construct the BSC (Kaplan and Norton, 1992, 1993). In later publications, they seem to change this opinion since they also discuss the use of personal BSC as an input for the annual bonus (Kaplan and Norton, 2000).

Table 1.2: Performance models

Performance models ¹	Definition of performance
Maximizing model	Long-term value of company (stock returns with efficient market)
Political model	Dominant coalition defines performance
Constituency model	Different coalitions (e.g., workers, shareholders) define performance
Business model	Overall performance improves when individual measures improve

Notes:

¹ Source: Meyer and Gupta (1994).

In the empirical analysis in this dissertation elements of both the political and the business model are used.¹⁷ The political model assumes that the dominant coalition of the organization chooses the performance measures that should be maximized. The difference with the maximizing model is that in the political model more than one measure is optimized, while in the maximizing model only the long-term value of the organization is maximized, operationalized by the current share price in an efficient market (Meyer and Gupta, 1994). The business model is a heuristic that helps to identify causal links between different dimensions of performance. These different dimensions of performance are often expected to be correlated in lagged periods rather than in the same period. The business model assumes that an increase in one measure leads to an improvement of overall performance (Meyer and Gupta, 1994).

In this dissertation the change in the PMS is evaluated through the items of the managers' contracts. These contracts are contrived by the dominant coalition of the organization, i.e., the top management, and should represent causal links between the dominant strategic drivers for the performance of the areas. The second research question assesses the change in the PMS. I assume that overall performance increases when each individual measure increases, i.e., I use the business model.

1.5 Contributions of the study

The research stream on which I build my empirical studies in chapter 5 and 6 is characterized by a lack of a consistent body of theory. The so-called “practice defined” variables financial and non-financial performance measures are attributed a number of characteristics without clearly understanding its properties (Luft and Shields, 2001). Only a number of ad hoc claims in front of these variables are available in the literature. This dissertation therefore mainly contributes to the empirical literature that addresses these claims. Although this might be a weakness of the motivation of the two empirical studies, it is also the only way forward to be able to improve the topic. The directions of further research in section 7.6. give a number of questions that should be answered before it is possible to be able to integrate the different findings in the literature.

This dissertation makes a number of contributions to the literature. First, it assesses the relative information content of non-financial performance measures and financial performance measures for future financial performance. This is the first direct test of

¹⁷ In contrast, the principal-agent model used to derive the expectations for the second research question uses the maximizing model (see section 6.2).

the popular claim that non-financial measures are better indicators for future financial performance than lagged financial measures. In addition, it adds to the growing literature of incremental information content of non-financial performance measures beyond financial performance measures for future financial performance (Ittner and Larcker, 1998a, Banker et al., 2000b, Nagar and Rajan, 2001).

This study contributes indirectly to the principal-agent literature. The relative information analysis addresses the congruence of the different measures with future financial performance, in which the congruence of the measure determines partly the weight of the measures in the contract (Datar et al., 2001). The incremental information analysis assesses the informativeness of each measure, i.e., whether the measures should have a non-zero weight in the contract (Ittner et al., 1997).

A lag search procedure, which is widely used in economic research, but relatively new in management accounting research, is used to explore the incremental information content of lagged non-financial measures beyond financial measures. The method used in my dissertation extends the method used in Banker et al. (2000b) in two ways. First, the various non-financial measures can have different lags for different financial measures. In the study of Banker et al. (2000b), the two different non-financial measures have the same lag for cost, revenues and profit.

Second, in Banker et al. (2000b) the average of the lagged non-financial measures from each period t is used to measure the lagged non-financial construct. This assumes that the lagged non-financial measure from each period t has the same impact on the financial measure. This assumption is unlikely to be met, since non-financial measures in period t might not have an immediate impact on the following period financial measure. In addition, later periods probably have a decaying effect on the financial measures. Therefore, in this study the factor loadings are also used to estimate the impact of non-financial measures on financial measures.

The evaluation question contributes to the empirical management accounting literature by exploring the long-term impact of a change in the PMS in a large Dutch service organization. Before the change managers were evaluated mainly on financial performance measures, whereas after the change worker satisfaction and customer satisfaction measures were added to these contracts. Exploring the long-term impact of such a change in the PMS is mentioned as a fruitful avenue for further research by Perera et al. (1997), and is for the first time considered in Banker et al. (2000b). Since non-financial measures are argued to have a long-term orientation, a cross-sectional study that measures the effect on only a short-term success measure will probably not capture the complete effect (Perera et al., 1997).

The evaluation question contributes indirectly to the multi-task principal-agent model by addressing the value of additional measures in the contracts of the managers.

The second study extends the study of Banker et al. (2000b) on a number of aspects. First, in the study of Banker et al. (2000b) two changes took place simultaneously. Next to the added measures in the contract the bonus of the managers was raised as well. This makes it hard to isolate the individual effects. In my study only one change took place, i.e., the contracts of the managers were added with the non-financial measures worker satisfaction and customer satisfaction.

A final contribution is that the research setting where the analyses are accomplished is different from the setting in Banker et al. (2000b) in a number of ways. First, the reward structure of the managers is different, i.e., the bonus is less important whereas promotion opportunities are more important. Second, in the research setting of Banker et al. (2000b) customer satisfaction is the most important

strategic issue, whereas in my research setting a quality orientation is the most important issue. Third, as a consequence from the second point the non-financial measures considered were different than in other studies.

1.6 Summary and overview of remainder dissertation

In this chapter the basic research questions and its motivation are outlined. Section 1.2 discusses the financial and non-financial performance dimension, and enumerates claimed advantages of these measures. Afterwards, section 1.3 positions the research questions in the management control literature. Next, section 1.4 explains the dependent variable in the evaluation question, i.e., managerial performance. Finally, Section 1.5 enumerates the contributions of the study.

The remainder of this dissertation is structured as follows. In chapter two the literature with respect to the two research questions is reviewed. This chapter describes the literature, gives a critical discussion, and provides directions for further research. Chapter three gives background information about the company where the data-collection took place. I discuss the organization, its PMS and the change and goals of the change in the PMS. This background information is given to facilitate the interpretation of the empirical results in chapter 5 and 6. In Chapter 4 I define all variables used in the two empirical studies and discuss the data collection process. Further, the chapter discusses properties of pooled time-series data. For the empirical analysis a pooled time-series data set is used. Pooled time-series data have unique econometric properties that have both advantages and disadvantages. Finally, I give the data description.

The empirical part of the dissertation is organized in Chapter 5 and 6. I divide the empirical work in two chapters since both analyses used a different framework and different research methods. Chapter 5 addresses the first research question, i.e., whether non-financial performance measures have relative or incremental information content, or both, beyond financial performance measures in predicting future financial performance. In this chapter I provide the framework that guides the research question, followed by the research methods and specifications used for the analysis. Then I give the empirical results. Chapter 5 ends with a discussion and summary. Chapter 6 addresses the second research question, i.e., whether managers perform better when they are evaluated relatively more on non-financial measures. This chapter starts with the framework that guides the research question. Afterwards, I discuss the research methods used for the analysis. Then the empirical results are given. Chapter 7 summarizes and discusses the findings in the dissertation.

Chapter 2

Literature review

2.1 Introduction

The purpose of this chapter is to review the available empirical studies, that explore relationships between financial and non-financial performance measures, that identify possible antecedents for non-financial performance measures use, and that consider contingencies where this use might be effective. Further, directions for further research are given to improve this research.

The deficits of using only accounting performance measures for evaluating managers and organization (-units) were recognized decades ago. In early research, most attention was paid to the consequences of, and identify circumstances in which these deficits were the most severe. This literature became known as the 'Reliance on accounting performance measures'-debate. For a recent overview of this literature, see Hartmann (2000). After the mid 80's, specially after the publication of "Relevance lost" (Johnson and Kaplan, 1987), more attention was paid to using other performance measures in addition to accounting measures. One example of these measures is non-financial performance measures, the subject of this dissertation.

While the advantages of the use of non-financial performance measures are well documented (see section 1.2), there are few papers that identify determinants of non-financial performance measures use. Even fewer studies measure outcome variables, either financial, non-financial, or at the individual level, when the appropriate match between non-financial performance measurement use and contingency variables exists.

Next to the question whether using more non-financial measures has positive performance effects, another important issue is the relationship between financial and non-financial performance measures. These studies consider whether non-financial performance measures are indicators for financial measures, what the time lag between improvements on non-financial measures and improvements on financial measures is, and the linearity of these relationships. The main motivation for this stream of research is the notion that non-financial measures are indicators for financial performance measures (Kaplan and Norton, 1992). These relationships can

be either in the same period or have a time lag before higher non-financial measures lead to higher financial performance.

Section 1.3 discussed the financial and non-financial performance dimension. The papers reviewed, however, use different labels, for example manufacturing performance measures, non-traditional information and reward systems, and operation-based non-financial measures. Therefore, the predominant reason to incorporate a study in this review is the content of the measure.

The chapter is structured as follows. The next section gives an introduction to the studies that address the subject of non-financial performance measurement. Section 2.3.1 provides an overview of studies that examine relationships between financial and non-financial performance measures. In addition, section 2.3.2 discusses some characteristics of these relationships. Studies that consider antecedents for the relative use of non-financial versus financial performance measures are reviewed in section 2.4. The performance effects of non-financial performance measure use are discussed in section 2.5. After this overview, section 2.6 discusses the literature. Finally, avenues for further research are indicated.

2.2 Overview empirical research

Papers for the review were selected based on a manual search in major accounting journals. A manual search was preferred above an electronic search, for example, on the words “non-financial”, because this would not be complete due to the various labels in the literature for the financial and non-financial dimension (see section 1.2). Further, articles were traced by references. Finally, only studies that analyze non-financial performance measures for internal reporting purposes are selected.¹

The first part, section 2.3, reviews studies that examine relationships between financial and non-financial measures of performance. One of the assumptions of the Balanced Scorecard is that non-financial measures are leading indicators of financial performance (Kaplan and Norton, 1992). These studies test this assumption. The potential time lag before the “lag” measures influence the “lead” measures is also explored.

For the non-financial measures there is a bias toward customer satisfaction measures. The majority of the studies reviewed are industry or organization specific. One motivation for this is that the relevance of specific non-financial performance measures, is different across industries (Amir and Lev, 1996). Another more pragmatic reason is that archival data, the primary datasource used for these studies, for non-financial measures from different organizations with the same variable definition is difficult to gather.

The second part, section 2.4 and 2.5, reviews studies that examine factors that drive the relative use of, or the emphasis placed on, financial versus non-financial performance measures. Some of these studies also assess the usefulness of the hypothesized match between the contingency variable and non-financial performance measure use. Both use and usefulness can be considered as success measures (Foster and Swenson, 1997).

At least two theoretical streams can give insights in the usefulness of using additional non-financial performance measures. First, the multi-task principal-agent theory states that additional performance measures can have value in the case where the “old” measure is incongruent with the goals of the principal and has noise

¹ For example, other studies analyze non-financial measures for predicting failure of companies (Keasey and Watson, 1987) and its value relevance (Lev and Thegaranjan, 1993).

(Feltham and Xie, 1994). This is a universal theory, i.e., more measures can never be worse than fewer measures.² Second, Milgrom and Roberts' (1995) notion of complementarity between operational strategy and the design of the management accounting system is explored.³ This assumes that a particular strategy, for example a quality strategy, can only be successful when other organizational systems, for example the performance measurement system and allocation of decision rights, are aligned with this strategy (Wruck and Jensen, 1994).

2.3 Relationships between financial and non-financial performance measures

2.3.1 Empirical evidence of relationships between financial and non-financial performance measures

One of the first studies that tested the relationship between a non-financial measure and financial performance is Anderson et al. (1994). They find evidence that a higher customer satisfaction leads to a higher return-on-investment (ROI) in both level and first-difference models.⁴ Customer satisfaction was measured from a survey of customers of 77 companies. Anderson et al. (1994) use the ROI measure because they consider this a measure for long-term economic health.

Banker et al. (2000b) use a longitudinal research methodology to better understand the relationship between non-financial and financial performance measures. This was an avenue for further research given by Chenhall (1997) and Perera et al. (1997). As non-financial performance measures they use two indicators for customer satisfaction, i.e, the likelihood of return of customers, and the number of complaints. The data set used is a pooled time-series of six years of monthly data from eighteen hotels from one hotel chain.

They find that one of the non-financial performance measures is associated with future operational profit and revenues. There was no relationship with the current period measures. In contrast to their expectations, costs⁵ did not increase significantly when the non-financial measures were higher. Costs were expected to increase because making customers happy will cost more, for example, due to hiring extra people to improve the service provided to customers. The results were consistent both in the level and percentage change specifications. Nagar and Rajan (2001) give as an explanation for the weak link between the complain measure in Banker et al (2000b) and financial measures that complain measures are not very good indicators of customer satisfaction.

Behn and Riley (1999) make another important contribution. They explore the impact of the non-financial measures customer satisfaction, market share, available ton-miles, and load factor on financial performance. They use eight years quarterly data of seven U.S airline companies. They find that non-financial performance measures are associated with current operating income, revenues and operating expenses. In addition, non-financial measures in the first and second month of a

² The result that additional measures can never have negative value is driven by the assumption that contracts are costless (see section 6.2.5).

³ These so-called Edgeworth complements assume that doing more of one increases the returns of the others. Interestingly, in this view interactions between contingency variables and non-financial use can never be dis-ordinal.

⁴ A complete cause-effect chain is tested from expectations of quality, to perceived quality, to customer satisfaction, to ROI. The other relations are not discussed in this chapter.

⁵ In the cost measure extra bonus payments due to increased performance are incorporated, therefore making the analysis a cost-benefit analysis.

quarter are found to be good predictors for quarterly revenues, expenses and operating income. Because the non-financial measures are available earlier than accounting data, Behn and Riley (1999) argued that financial analysts can make better predictions when this information is disclosed. In this study the non-financial measures are not only good signals for future performance, but also for current performance.

Ittner and Larcker (1998a) also examine whether customer satisfaction is an indicator for financial performance. This relationship is tested at three levels, that is, at the individual customer, the business unit, and the organizational level.

Although the explanatory power for the models is low, higher customer satisfaction led to a higher retention-rate of customers, higher revenues per customer, and higher revenue changes per customer. This is tested in a sample of 450.000 customers of a large telecommunication company.

The analysis at the business unit level, whether higher customer satisfaction leads to higher future profit, is a cost benefit analysis. The relationship between customer satisfaction and different financial measures of the units gave mixed results, in a sample of 73 bank branches from one organization. In the level model, the customer satisfaction measure had a positive impact on future revenues, but the impact on expenses, margins and return-on-sales (ROS) was not significant. Higher satisfaction levels lead to higher future improvements in margins and ROS, but had no impact on future revenue increases and future changes in expenses. Finally, changes in customer satisfaction had no impact on all four financial measures used.

Finally, Ittner and Larcker (1998a) provide an analysis on the firm level. They show that releasing customer satisfaction information in the magazine *Fortune* had a positive impact on abnormal stock returns. This implies that disclosing this information is of value to the stock market.

The third analysis of Ittner and Larcker considers the value relevance of non-financial measures. One concern of evaluating managers on non-financial measures is the persuasiveness of the capital market that inhibits its use. Amir and Lev's (1996) study relaxes this concern for industries that never made any profit.⁶ They show that investors are relying heavily on non-financial information for the valuation of securities in their sample of 14 wireless telecommunication companies. The fact that wireless telephone companies are in a growth market and need a lot of investments in R & D and franchise development explains this result. Further, in the growth stage of the product life cycle accounting performance measures are more influenced by accounting rules like the treatment of goodwill (getting licenses in Amir and Lev (1996) example). Therefore, investors are more concerned about future profit potential (more subscribers, a higher penetration rate, etc.), than the negative current financial results. The financial measures individually⁷ did not have any value relevance for the securities, they were only relevant in combination with non-financial performance measures.

Nagar and Rajan (2001) document the link between both financial and non-financial quality measures and the impact on future sales. They show in a sample of 11 plants of a *Fortune* 500 firm, that the changes in defect rate, and on-time delivery are related to future revenues. Only the impact on future revenues is estimated,

⁶ In this (type of) industry it is possible to evaluate managers for a large part on non-financial performance measures. In this way, congruence between the goals of the organization (shareholders) and the management is higher. Of the ten cellular companies who disclosed information about managers' compensation, six used exclusively non-financial measures and four used a combination of financial and non-financial measures.

⁷ (Change in) quarter earnings per share and book value per share.

therefore it is not clear whether improving the non-financial measures is beneficial for the companies, i.e., leads to a higher profit.

In sum, all studies show that non-financial performance measures can be indicators for future or current financial performance.

2.3.2 Characteristics of the relationship between financial and non-financial measures

The former section discussed empirical evidence of the relationship between financial and non-financial measures. This section explores characteristics of this relationship. It discusses the lag length, linearity of the relationship, and whether the lagged financial measure is used in the specification.

Although leading and lagging performance measures are important concepts in the BSC, there is no formal theory about the length of the time lag before an increase in non-financial measures leads to improved financial measures. Therefore, most studies use an arbitrary lag in their specification. In Anderson et al. (1994), the customer satisfaction measure is gathered in the first half of the year, and the ROI measure is from the end of the year. That implies they have a six month time lag in their design. Ittner and Larcker (1998a) also use an arbitrary time lag of six months between customer satisfaction and future financial performance.⁸ Finally, the lag in Nagar and Rajan (2001) between changes in defect rate, on-time delivery and future revenues is only one quarter. Banker et al. (2000b) explore the lag length through a lag search procedure from the data. In their case the time lag was 6 months, meaning that non-financial measures in period 1 have predictive power for the financial measure until period 7.

Another difference between the studies is the assumed form of the relationship. Anderson et al. (1994) use the logarithm of the financial and customer satisfaction measure in the specification. This implicitly means that the relationship between the two measures is assumed to be non-linear. Ittner and Larcker (1998a) explicitly consider the form of the relation. They find that before an increase in customer satisfaction is beneficial a threshold value has to be reached. In addition, benefits of improved satisfaction vanish when customer satisfaction is already at a high level.

A final distinction between the studies is that most studies use the lagged financial measure as an independent variable in the specification (Anderson et al., 1994; Ittner and Larcker 1998a; Banker et al., 2000b; Nagar and Rajan, 2001), whereas others do not (Behn and Riley, 1999). This difference in specification leads to a different implication. Using the lagged financial measures as an independent variable in the specification assesses whether the non-financial measures have additional value beyond the lagged financial measure. Omitting this variable assesses whether the non-financial measure predicts the financial measure.

In sum, all studies use different specifications, which seriously inhibits the comparison of the results. Table 2.1 gives an overview of these studies and provides information about the authors, variables measured, and the samples used.

⁸ It remains unclear why Ittner and Larcker (1998a) use the average of the third and fourth quarter, and compare this with the average of the first and second quarter, instead of using the individual quarters. Further, they use no control for seasonal effects.

Table 2.1: Overview of studies that examine relationships between financial and non-financial performance measures

Author(s)	Non-financial measure(s)	Financial measure(s)	Sample ^a
Anderson et al. (1994)	Customer satisfaction	ROI	77 large companies
Amir and Lev (1996)	Number of subscribers and penetration rate	Return and price of securities	14 cellular telephone companies
Behn and Riley (1999)	Customer satisfaction, market share, load factor and, available ton-miles	Operational cost, revenues, and profit	7 Airline industry
Ittner and Larcker (1998a)	Customer satisfaction	Revenue, accounting performance, stock performance	73 branch banks
Banker et al. (2000b)	Customer satisfaction	Operational cost revenues, and profit	18 hotels from a hotel chain
Nagar and Rajan (2001)	Quality measures	Future revenues	11 manufacturing companies

Notes:

The table gives the author(s), non-financial and financial measures, and the sample used in the studies.

^aWhere the study of Anderson et al. (1994) is cross-sectional and the other studies use pooled time-series data.

2.4 Antecedents of non-financial performance measures use

2.4.1 Introduction

Of all variables that are antecedents for the use of non-financial performance measures, or moderators for the relationship between use of non-financial measures and performance, operational and organizational strategy are explored in more than one study. These factors are discussed in section 2.4.2 and 2.4.3. Next, variables that are examined in only one study are discussed in section 2.4.4.

2.4.2 Operational strategy

Most studies use dimensions of operational strategy as a determinant of the use of non-financial performance measures. This is not surprising, since the use of more non-financial performance measures is given as part of the solution to the ‘irrelevant’ state of management accounting practices (Johnson and Kaplan, 1987). One of the reasons given for this ‘irrelevance’ are the changes in the internal and external environment of organizations. As a reaction to these changes organizations have implemented new manufacturing practices such as Just-In-Time (JIT), flexible production automatization, Total Quality Management (TQM), continuous improvement, and workgroups. Success stories from Japanese companies about new manufacturing practices led to a series of studies about these practices and their impact on the control system. Operational strategy is in this chapter used as a ‘bucket term’ for these practices.

Abernethy and Lillis (1995) were one of the firsts that examined the relationship between operational strategy and the use of financial and non-financial performance measures. They found support for the hypothesis that organizations that use a manufacturing flexibility strategy use less efficiency-based performance measures, in which efficiency-based measures are operationalized as cost efficiency.⁹ They found their results through semi-structured interviews with managers from 42 organizations.

Perera et al. (1997) extend the research of Abernethy and Lillis (1995) in a number of ways. First, Perera et al. (1997) use a survey of a randomly chosen sample, instead of the semi-structured interviews in pre-selected industries. Second, Abernethy and Lillis (1995) tested only one dimension of a customer focused manufacturing strategy, i.e., manufacturing flexibility. Perera et al. (1997) identify four dimensions of a customer focused manufacturing strategy, that is cost, quality, flexibility and the dependability of supply. The dimensions of this customer focused manufacturing strategy are proxied by advanced manufacturing technology (AMT) and advanced management practices (AMP). Third, Abernethy and Lillis (1995) examined only efficiency-based measures. In contrast, Perera et al. (1997) use an instrument with a number of financial and non-financial measures. Thus, Perera et al. (1997) study the relationship between a customer focused manufacturing strategy and the use of non-financial performance measures in a survey of 105 companies.

They find that organizations that have a customer focused manufacturing strategy, that is score high on both AMT and AMP, use relatively more non-financial measures. AMP seems to have a stronger impact than AMT.

Although Chenhall (1997) is not explicitly considering antecedents of non-financial performance measures use, his data indicates a univariate relationship between using TQM strategies and emphasizing manufacturing performance measures.

Ittner and Larcker (1995) explore the relationship between advanced manufacturing practices, and non-traditional information and reward systems.¹⁰ As a proxy for advanced manufacturing practices they use TQM, because this is argued to be the starting point of all operational strategies.¹¹ Traditional managerial accounting systems are defined as systems that provide aggregated financial information, operational control is based on variances from budgeted standards, and the reward system is built upon financial performance measures. In contrast, a non-traditional system gives more timely physical measures of operational performance, gives information for causes of differences with standards, and the rewards are based upon more non-financial performance measures.

They find empirical support for the hypothesis that organizations that have a TQM-strategy use more non-traditional information and reward systems, compared to organizations that do not have this strategy.

The study of Ittner and Larcker (1997) is similar to their study of 1995, and uses the same sample of 249 surveyed companies. They examine the relationships between a quality oriented strategy, and the use of strategic control practices. The strategic

⁹Next to the relationship between performance measures and flexibility they also examine the relationship between manufacturing flexibility and integrative liaisons. This factor had a stronger effect on the use of non-financial measures than the design of the performance measurement system.

¹⁰ Results from both studies of Ittner and Larcker (1995, 1997) should be used carefully. Especially the financial/non-financial dimension in the second paper is only one aspect of strategic control systems considered.

¹¹ If this is the case, TQM is an antecedent of other operational strategies. This has an impact on the specification for other studies, that is TQM can have either a direct impact on the use of non-financial measures or an indirect effect via operational strategies. For example, in Perera et al. (1997) AMT and AMP are both determinants of the use of non-financial performances.

control practices have to supplement the financial controls such as budgets, and short-term profits. The hypothesis that organizations placing a greater emphasis on quality in their strategy also use more quality related strategic control practices was supported.

The results of Ittner and Larcker (1995, 1997) are corroborated in van der Stede et al. (2001). In a sample of 128 US and Belgian managers they also find evidence that companies with a quality based manufacturing strategy use non-financial measures more extensively. Surprisingly, companies also use the financial measures more extensively when using a quality oriented manufacturing strategy.

Carr et al. (1997) examine differences between 65 ISO accredited and 42 non-ISO accredited companies in reporting quality in physical (and financial terms) and traditional performance measures.¹² Some support was found that ISO-accredited organizations use more physical (non-financial) measures. No support was found that ISO accredited organizations use more financial quality measures, and use less traditional performance measures, where traditional performance measurement is, for example, variance analyses, and financial quality measures are, for example, cost of quality. Less traditional performance measures were expected because they are less useful with strategies that have long-term benefits. The results suggest that traditional performance measures and physical (non-financial) measures are used as complements instead of supplements.

Banker et al. (1993) examine the relationship between reporting manufacturing performance measures (MPM) to workers while implementing practices like TQM, JIT, and teamwork. The underlying motivation for this hypothesis is that when workers are encouraged to find ways of improving processes, productivity and products, they need information to do so.

In a sample of 362 shopfloor workers, they find empirical support that organizations that implement TQM, JIT, and teamwork report more manufacturing performance measures, about quality and productivity, to their shopfloor workers. More specific, providing workers with charts about defects, schedule compliance, and machine breakdowns are fruitful ways of reporting when TQM, JIT and teamwork strategies are implemented. Using the shopfloor workers as unit of analysis instead of middle managers or CEO's is a distinguishing characteristic of this study.

In the study of Ittner et al. (1997) of determinants of the relative weight placed on financial and non-financial measures in CEO's bonus contracts, companies that emphasize a total quality strategy placed a higher relative weight on non-financial measures. These results are from a survey of 317 firms.

Summarizing the studies, there is a substantive body of evidence that suggests that organizations that have operational strategies such as TQM, JIT or a customer focused strategy, use relatively more non-financial performance measures for control or reward purposes.

2.4.3 Organizational strategy

Three studies investigate organizational strategy as an antecedent of non-financial measure use. From a sample of 317 organizations, Ittner et al. (1997) consider determinants of the relative weight of financial and non-financial measures in the

¹² First, a manipulation check was done that ISO-accredited companies really had a quality focused strategy and better developed quality practices. This test was executed because organizations can also strive for ISO accreditation only for the purpose of certification signaling to customers. For 7 of the 8 aspects ISO accredited companies used these practices more.

CEO's contract. They find that non-financial measures had a higher weight for organizations with a prospect strategy compared with a defender strategy. The explanation for this is that organizations with a prospect strategy focus on new product/service market opportunities, and quickly need to adapt to changes in the environment. This strategy may need a considerable time to turn into improved financial results. Therefore, short-run financial measures might not capture all these effects in this situation. In contrast, a defender strategy focuses on operating efficiency and therefore short-run financial measures may be suitable.

In contrast, Van der Stede et al. (2001) find hardly any support for a correlation between organizational strategy, i.e., on the differentiation versus cost-leadership distinction, and increased use of non-financial measures in a survey of 128 Belgian and US managers. Only the customer-oriented non-financial measures were used more in a differentiation strategy.

Morisette (1998) also finds no relationship between the mix of information that managers use, that is financial versus non-financial information, and the strategy of the organization, in which strategy is measured by the prospect versus defender continuum (Miles and Snow, 1978). The data in this study is gathered by way of direct observation, archival data, and questionnaires and interviews of 42 managers from production, marketing and human resource functions. This data is collected from three manufacturing and three service organizations.

In sum, one study that uses organizational strategy as an antecedent for non-financial performance measures use finds a relationship (Ittner et al., 1997). In contrast, Morisette (1998) and Van der Stede et al. (2001) find no relationship between organizational strategy and the relative use of non-financial performance measures.

2.4.4 Other antecedents

A number of additional variables are investigated that can be determinants of non-financial performance measures use. Ittner et al. (1997) review determinants that drive the relative share of financial and non-financial indicators in the incentive scheme of CEO's. This study concentrates specifically on the formal compensation scheme. The focus of the study is on annual formula-based bonus plans, because these over-emphasize short-term accounting returns and discourage long-term investments (Kaplan and Norton, 1992).

They find that when financial measures have more noise, that is the performance measure is not a good indicator for the effort of managers, non-financial measures have a relatively higher weight. There are no relationships found with the level of financial distress, the value of CEO's equity holdings relative to salary and bonus, and the influence of the CEO over the board of directors. This last factor was motivated by Eccles and Mavrinac's (1995) study. They found that investors and market analysts believe that reported non-financial measures may be biased, and its computation may change over time, allowing the measure to be more easily manipulated. So when CEO's have more influence over the board they try to be evaluated more on non-financial measures.

Morisette's (1998) explanatory study identifies factors that influence the proportion of financial and non-financial information in the managers' mix of information used for performance monitoring. Empirical support is provided for the propositions that managers use relatively more non-financial performance measures in their information mix when managers perceive their task-technology as routine, perceive

they are evaluated by their superior on non-financial measures, and have more experience. The assumption that experienced workers have a better understanding of a few critical causal relationships between their work and performance explains this last relationship. Their focus is therefore on the non-financial measures of these relationships.

In contrast, relatively more financial information is used when managers perceive their task-technology as non-routine¹³, perceive they are evaluated by their superior on financial measures, and have less experience. No relationship was found with the perceived level of external environmental uncertainty or the level of managers¹⁴, i.e., strategic, tactical or operational. The most important factor found by Morissette (1998), but not used as a prior proposition in the study, was the difference between a throughput function versus an output function. Managers in an output (throughput) function all used relatively more financial (non-financial) performance measures. This result is also found in Bruns and McKinnon's (1993) field study. They find that manufacturing managers, i.e., a throughput function, use more operational non-financial measures, whereas marketing and sales managers, i.e., an output function, use more financial measures.

Schiff and Hoffman (1996) examine whether managers use different measures for the evaluation of managers or departments. This hypothesis is motivated by the controllability principle, that is not all factors which drive divisional performance can be controlled by managers. Therefore, the use of different performance measures for evaluating managers or departments is expected.

They did an experiment with 54 executives from a large retail organization. These managers were given 51 cases of 4 financial and 4 non-financial measures. Based on these measures they had to assess the overall performance of either managers or departments. They find the following results. First, executives used for the evaluation of departments and managers both financial and non-financial measures. Second, for the evaluation of managers, executives placed a higher weight on non-financial measures, while for the evaluation of departments the weight on financial measures was higher.

Coates et al. (1992) investigate the differences in use of performance measures between different countries. The data used are 45 interviews from 5 multinationals each in Germany, the United States, and the United Kingdom. Their conclusions are somewhat strange. First, companies in the US used more non-financial performance measures compared with their German and UK counterparts. Second, German multinationals used more market share and return on sales performance measures. Third, UK companies' performance measures were set out in more detail. The results are based on a small sample, and therefore, must be interpreted with caution. The results are explained by the influence of cultural and structural factors, such as the strong influence of the capital market in US and UK compared to Germany. However, one would expect that when the capital market has a stronger influence more financial performance measures would be used. In addition, differences between industries were found.

The studies in this section have an exploratory character. The variables are examined only once and therefore any conclusions should be derived cautiously. However, the studies seem to indicate that non-financial measures are used more often when financial measures have more noise (Ittner et al. 1997), managers are evaluated

¹³ Which was influenced by experience. Managers who were in their position for a short time saw their job as non-routine.

¹⁴ There was only an indirect relationship via the functional area of the manager.

instead of departments (Schiff and Hoffman, 1996), in throughput functions instead of output functions, when managers perceive they are evaluated on non-financial measures, managers see their task as non-routine, and managers have more experience (Morissette, 1998).

2.4.5 Summary

In sum, there is a considerable body of evidence that gives support for a number of antecedents for non-financial performance measures use. Comparing the results for the operational strategies (e.g., JIT, TQM) with the organizational strategies (e.g., build versus harvest) seems to suggest that the former has more impact on the use of non-financial measures. One possible explanation for this result is that the organizational strategy variables only have an impact on non-financial measures related to customers, whereas the operational strategies have an impact on all kinds of aspects of the organization (Van der Stede et al., 2001).

The studies are summarized in table 2.2.

Table 2.2: Overview of studies that examine antecedents of non-financial performance measures use

Author(s)	Antecedent	Financial/non-financial dimension	Unit of analysis
Abernethy and Lillis (1995)	Manufacturing flexibility strategy	Efficiency/flexibility measures	Middle-managers
Perera et al. (1997)	Customer focused manufacturing strategy	Financial /non-financial performance measures	Middle-managers
Chenhall (1997)	Customer focused manufacturing strategy	Manufacturing performance measures	Middle-managers
Ittner and Larcker (1995)	TQM-strategy	Information and reward system	Middle-managers
Ittner and Larcker (1997)	Quality oriented strategy	Strategic control systems	Middle-managers
Carr et al. (1997)	ISO accreditation	Quality controls	CEO's, Middle-managers
Banker et al. (1993)	TQM, JIT, and team work strategies	Reporting manufacturing performance measures to shopfloor workers	Shopfloor workers
Ittner et al. (1997)	TQM- strategy, prospect/defender strategy, noise in financial indicator, CEO's influence over the board, financial distress, CEO's value of equity holdings relative to salary and bonuses	Weight of financial non-financial measures in CEO's compensation scheme	CEO's

Table 2.2: Overview of studies that examine antecedents of non-financial performance measures use (continued)

Morissette (1998)	Prospect/defender strategy, Perceived external environmental uncertainty, experience, perceived evaluation, task-technology, level managers, use of information, throughput/output function	Use of financial and non-financial performance in information mix	Strategic, tactical, and operational managers
Schiff and Hoffman (1996)	Department/manager evaluation	Financial/non-financial measures	Department, middle-managers
Coates et al. (1992)	Countries	Financial/non-financial performance measures	Multi-Nationals
Van der Stede et al. (2001)	Low cost/differentiating strategy, quality based manufacturing strategy	Financial/non-financial performance measures	Middle-managers

Notes:

The table provides information about the author(s), the antecedent variable(s), the financial/non-financial dimension, and the unit of analysis.

2.5 Performance effects of non-financial performance measure use

A considerable number of studies explore the performance impact of emphasizing non-financial measures relatively more. These studies often use a contingency approach, i.e., they test the performance impact of a match between the contingency variable and the emphasis on non-financial measures.

Abernethy and Lillis (1995) find that organizations that emphasize non-financial measures relatively more and have a manufacturing flexibility strategy perform better than organizations that do not have this match. From the interviews they held an interesting point emerged, that is about the moment you measure performance. Some organizations only recently changed their performance measurement system (PMS), or were still in the process of doing so. Nevertheless, performance had improved for organizations with the appropriate match. This point will be discussed more extensively in the discussion section.

In contrast, Perera et al. (1997) do not find support for any performance effect when organizations have the appropriate match between a customer focused strategy and operation based non-financial performance measures.

Ittner and Larcker (1995) find little support for the hypothesis that organizations that have a match between a TQM-strategy and non-traditional information and reward systems perform better. They used return-on-assets and quality as performance measures. This performance hypothesis is also tested with a time lag in performance after implementation of TQM and information and reward systems. This is examined by using the dependent variable of performance from the current period and TQM and information and reward systems practices of three years ago. The highest performers in the sample were organizations that used both little traditional information and reward systems and little formal TQM-practices. This suggests that formal strategic control practices might have a negative impact on performance.

Ittner and Larcker's (1997) study gives the same result as their 1995 study with a slightly different concept, i.e., quality oriented strategy and strategic control practices.

A third study that finds a similar result is Van der Stede et al. (2001). In their sample companies that use a quality based manufacturing strategy, and place much emphasis on non-financial measures also perform worse than companies that only use a quality oriented strategy without a formal control system.¹⁵

Chenhall (1997) investigates whether the relationship between TQM and performance is influenced by the use of manufacturing performance measures (MPM).¹⁶ He motivates this hypothesis from the notion that many TQM-initiatives do not give the predicted effects for profitability improvements if complementary parts in the organization, such as the PMS, are not adapted. In a sample of 39 organizational units, supporting evidence is found for this conjecture.

Chenhall (1997) controls for implementation effects of the new PMS by selecting the sample from organizations that implemented the PMS at least three years ago (analogous to Simons (1987)). Further, the sample has no organizations that use AMT. Otherwise, the study could measure the effect from the study of Perera et al. (1997). Chenhall (1997) gives different directions for further research. First, it would be interesting to follow the development of TQM and MPM's over time. Second, he suggests that it will be important to study how non-financial indicators combine with traditional financial performance measures.

Sim and Killough (1998) assess the impact on different dimensions of performance of complementarities between TQM or JIT, and providing specific goals, performance measures, and incentives in line with these strategies. The subjects in this sample were 83 directors of manufacturing. First, they find that the interaction of TQM or JIT, and providing specific goals had a positive impact on quality and customer performance. Second, they find no support that the interaction between TQM or JIT and providing non-financial measures more frequently had a positive impact on quality and customer performance. Finally, they find that providing incentives in the case of a TQM or JIT strategy has a positive impact on quality and customer performance. In this study, performance is measured by less aggregated components of performance. It is unclear, however, whether the benefits are worth the additional costs of providing more information.

Govindarajan and Gupta (1985) examine the relationship between organizational strategy and the use of short-run versus long-run criteria in the determination of the incentive scheme for SBU managers. Organizational strategy is operationalized with the build or harvest continuum. Long-run criteria consist of the items sales growth, research and development, market share, new product development, market development, personal development, and political and public affairs. Short-run criteria consist of the items ROI, cost controls, cash flows, operating profits and profit margins. Thus, although they use the labels short-run and long-run criteria, the scales used consists out of financial and non-financial performance measures.

From a sample of 58 strategic business units they find empirical support that a greater reliance on long-run criteria for determination of the SBU general managers' bonuses is more effective (profitable) with a build strategy. Surprisingly, a greater reliance on short-run criteria in combination with a harvest strategy is not more

¹⁵ They find that companies that make extensive use of subjective measures and a quality based manufacturing strategy are the highest performers.

¹⁶ Although Chenhall answers the research question whether the relationship between TQM and performance is moderated by the reliance on MPM, the statistical method used also answers the question whether the relationship between the reliance on MPM and performance is moderated by TQM.

effective. The interpretation for these results seems to be that short-run measures are always important, while long-run criteria are only important in a build strategy.

Although being one of the first who examined the use of relatively more non-financial measures, they already recognized the caveats in this research, that are the difficulties of measuring performance (effectiveness) and the cross-sectional methodology.

In sum, the results for the performance hypothesis, when the appropriate match between operational strategies and non-financial performance measure exists are unclear. Although Abernethy and Lillis (1995), Chenhall (1997), and Sim and Killough (1998) present favorable results, Perera et al. (1997) and Ittner and Larcker (1995, 1997) find no effect.

The studies discussed above are all cross-sectional studies. Therefore, these studies are all subject to the same caveats. First, the outcome measures used to indicate the success of emphasizing non-financial measures relatively more are short-term outcome measures that do not incorporate the full performance effects (Perera et al., 1997). Second, since the studies use a cross-sectional methodology, performance is measured at one moment and therefore does not capture long-term performance effects of emphasizing non-financial performance measures (Perera et al., 1997).

These problems motivated Banker et al. (2000b) to consider performance effects of placing more emphasis on non-financial performance measures in the evaluation and reward system of managers in a longitudinal study. Pooled time-series of 6 years of monthly data from 18 hotels from one hotel chain were available for the analyses.¹⁷ From the quasi-experimental design, the impact of the change in the performance measurement system is examined on both financial and non-financial dimensions of performance. After including non-financial measures in the compensation contract of the managers, the customer satisfaction measures increased. Furthermore, the revenue and profit measures rose, when relatively more emphasis was placed on non-financial performance measures in the compensation contract. In contrast to their expectation, the cost measure was not rising. This was expected because the emphasis on the cost measures is decreasing after including the customer satisfaction measures in the contract, and in addition, increasing customer satisfaction may cost money.

The results for the performance hypothesis are summarized in table 2.3.

Table 2.3: Overview of studies that examine performance effects of non-financial performance measures use

Author(s)	Fit between contingency variable and non-financial performance measure use ¹	Performance hypothesis Confirmed?
Abernethy and Lillis (1995)	Manufacturing flexibility and manufacturing performance measures	Yes
Perera et al. (1997)	Customer focused strategy and operation based non-financial measures	No
Chenhall (1997)	TQM and manufacturing performance measures	Yes
Ittner and Larcker (1995)	TQM and non-traditional performance and reward systems	No

¹⁷ Limited control data was available from hotels of the same hotel chain that were franchised.

Table 2.3: Overview of studies that examine performance effects of non-financial performance measures use (*continued*)

Ittner and Larcker (1997)	Quality oriented strategies and strategic controls	No
Van der Stede et al. (2001)	Quality based manufacturing strategy and non-financial performance measure use	No
	Low cost/differentiation and non-financial performance measures use	No
Govindarjan and Gupta (1985)	Build/harvest and providing long-run/short-run in the bonus system	Yes
Sim and Killough (1998)	TQM/JIT and providing specific goals	Yes
	TQM/JIT and providing performance measures	No
	TQM/JIT and providing incentives	Yes
Banker et al. (2000b)	²	Yes

Notes:

The table provides information about the author(s), the financial/non-financial dimension, the antecedent variable(s), and whether there is support for the performance hypothesis.

¹ I do not distinguish between different types of “fit” (see Venkatraman, 1989).

² Banker et al. (2000b) do not consider moderators in the relationship.

2.6 Discussion and future research

2.6.1 Relationships between financial and non-financial performance measures

Anderson et al. (1994), Amir and Lev (1996), Behn and Riley (1997), Ittner and Larcker (1998a), Banker et al. (2000b) and Nagar and Rajan (2001) give support for the argument that non-financial performance measures can be leading indicators for dimension of financial performance. Table 2.1 summarizes these studies. The methods used in the studies differ, and the samples are from different industries. This raises a number of issues.

First, to explore relationships between financial and non-financial performance measures, and specially to find a time lag, time-series data are needed. Chances to find comparable time-series about performance measures for a large sample of companies are almost zero. However, an obvious disadvantage of the industry or organization specific samples used in the studies is the generalizability of the results to other samples. Another disadvantage of industry or organization specific studies is the ability to test contingency variables in the relationships. Industry or organization specific studies will probably show less variation in contingency variables, because organizations from one industry are on average more homogeneous. Therefore, the only solution seems to be to examine the relationships between financial and non-financial measures, and the time lag, with time-series data of one or a small set of organizations, and contingencies that influence these relationships or the time lag, in large sample (cross-sectional) studies.

Second, for the time lag between improved non-financial measures and improved financial measures little research is done. In a few studies an arbitrary time lag is used (Anderson et al., 1994; Ittner and Larcker, 1998a; Behn and Riley, 1999; Nagar and Rajan, 2001). The method used in Banker et al. (2000b) to explore the time lag from

the data is the first try to specify the lag from the data.¹⁸ The specification used to explore this time lag is (given without control variables):

$$F_t = \alpha + \beta_1 * F_{t-1} + \beta_2 * NF_{t-1} + \varepsilon_t \quad (2.1)$$

In which F_t is a financial measure and NF_{t-1} is the lagged non-financial measure. In this specification β_2 gives the impact of the lagged non-financial performance measure on future financial performance. The motivation to incorporate F_{t-1} in the specification is that if non-financial measures have no additional value over financial measures, only the financial measures can be used instead.

The time lag search is carried out by adding lagged non-financial measures, i.e., NF_{t-2} , NF_{t-3} , etc., in equation (2.1) until the “best model” is found. This time lag search assumes, however, that the impact of an increasing non-financial measure is immediately affecting the financial measure in period $t+1$. This is not necessarily congruent within the concepts of leading and lagging measures. It is equally likely that the non-financial measure at t has an impact on the financial measure not before for example $t+3$. Further, the time lag found in Banker et al. (2000b) between the two non-financial measures and all financial measures, i.e., cost, revenues, and profits, is six months. Whether this is a coincidence or a restriction in the analysis is not clear from this paper.

The different methods used in the studies to assess the lag between improved non-financial measures and improved financial measures means that they implicitly use different definitions. Most studies use an arbitrary lag and therefore implicitly assume that the impact of non-financial measures on future financial measures starts not before, for example, 6 months and that the complete impact of the improved non-financial measures materializes in this period. This assumption leads to an understatement of the impact of non-financial measures on future financial measures both when the actual lag is shorter and longer and therefore, I conjecture that this leads to measurement error and that the effects found are therefore smaller than the “real” effects. In contrast, the lag search procedure used in Banker et al. (2000b) assumes that the impact of non-financial measures has an immediate impact on future financial performance in the next period and that this impact vanishes after, for example, 6 months.

It is often assumed that non-financial measures are better indicators for future financial performance. The specification used in equation (2.1), however, is not testing this. Instead, it tests whether non-financial measures explain additional variation over the financial measure at $t-1$ of the financial measure, not necessarily more. Explaining additional variation is called the incremental information concept (Biddle et al., 1995). In contrast, the question whether non-financial measures are better indicators than lagged financial measures for future financial performance, is the relative information content. Related to this observation is the fact that value-relevance of non-financial performance measures, i.e., whether they explain more than the lagged dependent financial measure, is not the same as usefulness for control purposes. For example, it is often argued that non-financial measures are more informative about causes than financial measures are (Singleton-Green, 1993). Thus,

¹⁸ This results of course in a data driven procedure. Although it would be helpful to estimate the lag, and then use a holdout sample to validate this lag, this is a costly procedure. To assess the lag from the data, long time-series are needed, therefore the need to split this data leads to very long time-series.

providing non-financial measures might be more helpful for improving future financial performance than providing the lagged financial measures.

Although researchers are of course limited to work with data that is available, attention should be paid to the measurement of the non-financial measures. The non-financial measures are not always measured in a reliable and valid way in organizations. Therefore, care should be taken with using this data in analysis (Lambert, 1998).

The literature and the review given above suggest a number of fruitful avenues for further research. First, Because non-financial measures are said to pay more attention to the long-run perspective of the organization, it is expected that a time lag between improved non-financial measures and improved financial measures will exist. However, research that examines factors that determine this lag length, is still unavailable. Therefore, further studies that examine determinants of the time lag length are valuable. A number of possible factors could be the repurchase time of products, and the number of competitors. Additionally, the method to test this time lag needs further development.

Second, the form of the specification used to test relationships between financial and non-financial measures, i.e., either level or change models, is an often discussed issue in this research stream. An advantage of change models is that they control for omitted variables in the specification (Lambert, 1998). Banker et al. (2000b), however, argued that level models are more appropriate, because change models assume that the non-financial measures from different periods have a constant impact on the financial measures at period t .¹⁹ This assumption is clearly not valid. Next to the question what the right specification from an econometric viewpoint is, the different models have different theoretical implications and therefore are not additive support for one and the same hypothesis. This should be incorporated into theory building in further research.

A further almost unexplored area is the linearity of the relationships between financial and non-financial measures. Although in so-called Japanese management the zero deficits and free quality philosophy is often advocated (Juran, 1979), diminishing returns to scale between improved non-financial and financial measures are very likely. Ittner and Larcker's (1998a) study is the only study that examines this non-linear relationship.

Finally, the theoretical concepts of leading and lagging measures assume that non-financial measures are indicators for financial measures. However, improving non-financial measures probably will cost money first. Therefore, statistical methods that take the bi-directional causal nature of the relationship into account need to be used to explore relationships between financial and non-financial measures (Luft and Shields, forthcoming). This also suggests that the impact of non-financial measures on future financial performance is not the same for each dimension of financial performance, e.g., cost, revenues or profit.²⁰

¹⁹ Additional concerns of using change variables are its lower reliability, and dependence on the level of the measures (Cohen and Cohen, 1983).

²⁰ In a cross-sectional data set from more than 1000 retailers, Banker and Mashruwala (2002) find simultaneity between customer satisfaction and revenues. They conjecture that higher customer satisfaction does not lead to higher revenues, but that higher revenues leads to more traffic in shops and therefore to lower customer satisfaction.

2.6.2 Use and usefulness of non-financial performance measures

The studies discussed in section 2.4 explore a large number of antecedents of non-financial performance measures use, especially operational and organizational strategies. Whether organizations or organizational units perform better when they have the proper match between non-financial measure use and the contingency factor is unclear. For example, in Ittner and Larcker (1995, 1997) and Van der Stede et al. (2001), the highest performing companies were the companies that had a low emphasis on using non-financial measures, but had a quality oriented strategy.²¹ Carr et al. (1997) also argued that using a TQM-philosophy might be just as successful compared to a formal system. This suggests that the use of formal controls might inhibit increased performance.

Govindarajan and Gupta (1985), Carr et al. (1997), and Van der Stede et al. (2001) find that non-financial measures are used more with a build strategy, in ISO-accredited organizations, or with a quality based manufacturing strategy, but that financial measures are not used less in these situations. This suggests that financial measures are always important but that non-financial measures are important only in certain circumstances.²²

The studies that identify determinants of the relative use of non-financial performance measures show a heterogeneous picture. The studies seem to be an example of the disappointing state of nature of contingency theory of management accounting as shown in Otley (1980). One variable is never used twice in different studies, even by the same authors (for example, see Ittner and Larcker, 1995, 1997).

The measurement of effectiveness in contingency theory is a problem in general (Otley, 1980). There are several aspects of this problem specific for this subject. First, because non-financial measures are said to place more emphasis on the long-run position of the organization, long-run impacts can be expected. However, the performance measure used to assess effectiveness is often measured at one point in time, and therefore does not capture this effect. Thus, while the problems of an appropriate outcome variable are a general concern, it is even bigger in the area of non-financial performance measures use. Ittner and Larcker (1995, 1997) are partly controlling for this effect by explaining current performance by lagged practices. This also mitigates effects of implementation, i.e., the companies are in different states of implementation.

A second problem is that programs, such as TQM and investments in more new manufacturing practices, are more often implemented at moments that organizations are facing bad times instead of being on track. This will influence the perceived usefulness of the change in the performance measurement system. Thus, the moment of measuring effectiveness is important (Abernethy and Lillis, 1995).

Next to the moment of measuring, the instruments used to measure performance are different in each study. This further inhibits the comparability of the results. First, Sim and Killough (1998) assess performance on non-financial dimensions. Second, Chenhall (1997) and Perera et al. (1997) measure performance growth. Third, Ittner and Larcker (1995, 1997) use both return-on-assets, and a quality measure. Finally,

²¹ In contrast, Van der Stede et al. (2001) find that companies that have a quality based manufacturing strategy and use subjective performance measures are the highest performers.

²² This result implies that the total emphasis on performance measures is not constant. When in the case of a build strategy or with ISO-accredited companies financial measures keep the same emphasis and the emphasis on non-financial measures increases, this means that the total emphasis on performance measures increases.

Abernethy and Lillis (1993) compare performance of departments to other departments, but do not specify on which dimension(s) of performance this is done. Banker et al. (2000b) assess a change in the contract of managers, from using relatively more financial to using relatively more non-financial measures, on both financial and non-financial dimensions of performance. Otley (1980, p. 424) gives as advice for measuring effectiveness that:

“the evaluation of the appropriateness of particular varieties of accounting control systems must therefore take place by comparison with a range of measures of effectiveness, at both an organizational and individual level of analysis”.

However, he meant that a particular system should be evaluated on a number of different performance dimensions in one study, not that each study should pick its own measure.

Sim and Killough (1998) use an absolute level of performance²³, whereas Chenhall (1997) uses a relative measure, which is compared to the industry. Performance on a relative measure compared to the industry depends on practices of other companies in the industry, whereas an absolute measure does not. For example, when all companies in an industry use a TQM strategy in combination with the appropriate level of non-financial performance measures use, no performance effect is expected when a relative measure is used.

Assessing the impact of the use of non-financial performance measures on performance at the individual level is not done so far. A skillful borrowing from the ‘Reliance on accounting performance measures’-debate can be done, in which outcome variables like job-related-tension (JRT) and dysfunctional behavior are used.

The limitations of measuring performance in cross-sectional studies motivated Banker et al. (2000b) to assess performance effects of using more non-financial measures in the contracts of managers in a longitudinal study. This in itself valuable approach is problematic because it is difficult to get a control group for the units that use more non-financial measures. Further, it is difficult to test contingency measures in the relationships.

In general, the studies that measure performance hardly motivate the outcome measure used. Further research should make it explicit on which dimension of performance, i.e., individual, non-financial or financial, improvements are expected.

Another characteristic of these studies is the different measurement scales used for the financial versus non-financial construct. Abernethy and Lillis (1995) were one of the first to develop a measurement scale. Although others have created instruments based on this scale, they are all different. This is also a sign that the research area of non-financial measurement is relatively young. Two different methods are used to measure the relative use of financial and non-financial performance measures. First, several financial and non-financial performance measures are given. The respondents assess on a Likert-scale the use or emphasis of these performance measures in their organization (Govindarajan and Gupta, 1985; Abernethy and Lillis, 1995; Perera et al., 1997; Van der Stede et al., 2001). This method could be rationalized by the argument that non-financial performance measures often have a different meaning and relevance in specific industries, and therefore the scales should be adapted for each study. However, the diverse samples used in each individual study are not supporting this argument. The use of a common set of performance measures for companies from different industries might therefore influence the results. Second, a relative

²³ For example, one item is the percent of customer complaints.

measure of the financial versus non-financial construct, either a Likert or percentage scale, is used (Chenhall, 1997; Ittner et al., 1997; Morisette, 1998). The discussion above suggests that the second way might be a more precise scale, because this approach is more neutral to the different meaning of non-financial performance measures in different industries.²⁴

Summarizing the discussion about the non-financial versus financial dimension, further attention should be paid to develop the financial versus non-financial construct. For example, by considering the number of dimensions in the items used with factor analysis. In addition, the question whether the financial versus non-financial dimension has any discriminant validity from, for example, scales that measure accounting versus non-accounting should be answered.

A major econometric problem is that the contingency variables and the non-financial performance measurement use are both choice variables for the organization. Therefore, the independent variables are endogenous in the performance hypotheses. This leads to problems of endogeneity and simultaneity, which are difficult to solve (Ittner and Larcker, 1997, 2001).

Another difference between the studies is the purpose of using non-financial measures. It might make a difference whether non-financial measures are used for decision-making purposes or for decision control (Zimmerman, 1997). For example, in Govindarajan and Gupta (1985), Ittner et al. (1997), and Banker et al. (2000b) the non-financial measures are used for the bonus of the managers. In Banker et al. (1993), Chenhall (1997), and Perera et al. (1997) the measures are only provided to the managers to make better decisions. Perera et al. (1997) suggest that the advantages of the use of non-financial measures, in their case operation-based measures, are more motivational than instrumental.²⁵ This suggests that by giving managers specific goals they are only improving on these measures, whereas this is not necessary beneficial for the organization.²⁶ In contrast, in Sim and Killough (1998) the hypotheses related to giving specific performance goals, and contingent rewards, i.e., motivational variables, are significant, whereas only providing the measures is not.²⁷ Therefore, the purpose for using non-financial measures, for example for compensation or decision making, should get more attention in the theories underlying future studies.

With respect to the samples, in the studies that address contingencies mostly samples are used from manufacturing organizations (Banker et al., 1993; Abernethy and Lillis, 1995; Chenhall, 1997; Ittner and Larcker, 1995, 1997; Sim and Killough, 1998), as the available literature is of course biased by the 'operational strategy'-contingency factor studies. Relatively less attention is paid to service organizations.²⁸ Although there is no a priori reason that things might be different in such organizations, it would be interesting to test this.

²⁴ However, this approach has the problem that it leaves too much to the perception of respondents.

²⁵ Motivational means that managers have a specific goals which they try to maximize and instrumental is to help the managers implementing strategies by providing measures and giving feedback about these strategies.

²⁶ This not only suggest that the motivational premise is stronger than the instrumental, but also implicitly means that the non-financial measures are not good indicators for financial performance.

²⁷ It is interesting to compare this with the evaluation of ABC-systems. Anderson and Young (1999) find that providing rewards is one an important factor for ABC system evaluation.

²⁸ A notable exception is Morisette (1998), who examines three manufacturing and three service organizations, but finds no differences.

2.7 Summary and conclusions

This chapter reviews two related branches of literature about non-financial performance measurement. The first branch are papers that examine relationships between financial and non-financial performance measures. All studies suggest that the non-financial measures are indicators for either current or future financial performance. Little evidence is available about a possible time lag between leading, i.e., non-financial measures, and lagging measures, i.e., financial measures. Although premature, there is evidence that relationships between non-financial measures and financial measures might be non-linear.

Further research in this branch should pay attention to contingency factors that influence the relationships between specific non-financial and financial measures, and that influence the time lag. The specifications used for testing the relationships between non-financial measures and financial measures, and the possible time lags, i.e., equation (1.1), need further development.

The second part examined antecedents of non-financial measure use, and contingency factors that influence the relationship between non-financial performance measure use and performance. Of the antecedents of non-financial measure use there is consistent evidence that when organizations or organization units have operational strategies like TQM or JIT, non-financial measures are used or emphasized more. Also, when organizations have a build or prospector strategy more non-financial measures are used than when organizations have a harvest or defender strategy. For a number of other determinants tested, it is too early to derive conclusions. Results on effectiveness when the hypothesized match between non-financial performance measure use and the contingency factor exists are inconclusive. There are specific limitations with the studies that need attention in further research.

Especially, measuring performance needs further attention. In theory building it should be made clear which dimension(s) of performance, either financial, non-financial, or individual, are expected to increase when there is a match between non-financial measurement use and the contingency variable. This also makes it clear on which level, for example, at the business unit, organization, or capital market, the improved performance is expected. In addition, the moment of measuring is important. Because non-financial measures pay more attention to the long-term position of the organization a short-run measure might not capture the effect.

Second, the measurement for the construct financial versus non-financial measures needs to be developed further. It is not clear whether financial versus non-financial measures is an uni-dimensional construct, and has discriminant validity from, for example, accounting versus non-accounting measures.

Finally, since different studies consider different determinants of non-financial performance measures use, this might be a partly explanation for the different results found. Therefore, future studies should also indicate what the goal is for using the non-financial performance measures.

Chapter 3

Research site

3.1 Introduction

This chapter describes the research site where the empirical part of the dissertation took place. I give a description of the company, the PMS, the change in the PMS, the goals of this change, and the controllability of the different performance measures by the area managers. This is used as background information for the interpretation of the analysis.

The chapter is structured as follows. The next section describes the organization and gives the motivation of selecting the company for the empirical analysis. Section 3.3 discusses the change in the PMS that took place in 1996. Section 3.4 explains the goals of the change in the PMS. Section 3.5 outlines the complete performance measurement cycle. Section 3.6 describes the controllability of the different measures.

3.2 Organization

The empirical part of the study took place in the business unit Letters of PTT Post¹, the Dutch mail company. PTT Post is a daughter of KPN Holding and has approximately 55.000 employees. This company was privatized in 1989 and became listed in 1994 (Groot and Smidt, 2000). PTT Post delivers letters, parcels, international mail, direct marketing mail, non-addressed advertising mail and express mail to private persons and businesses (Groot and Smidt, 2000).

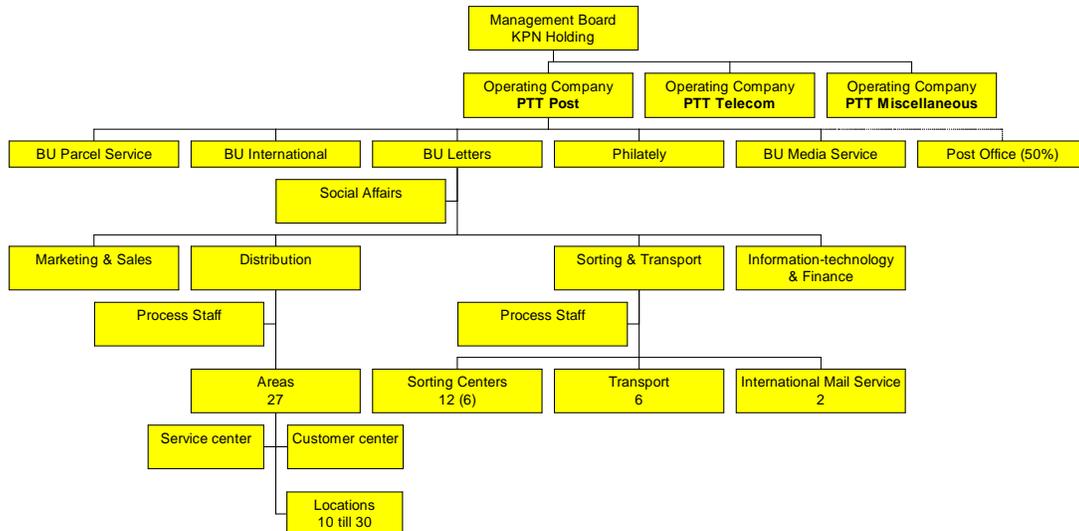
The primary task of the business unit Letters is delivering the different kind of postal items to the right customer at a pre-specified time period. It has a monopoly for letters up to 500 grams. An independent agency controls the price setting policy of the PTT Post to ensure that they are not misusing this monopoly. Before the on average 21 million postal items a day are delivered, the products are collected from the mailboxes and transported to the sorting centers. These sorting centers sort the postal

¹ Currently this is called TPG Post, but I will use the name of the company during the time period of the study. This chapter presents the company as it was during the period of study, i.e., between 1995 and 1997.

items and transport them to the areas. The areas sort the products again on zip code and distribute the postal items to the customers.

Figure 3.1 describes the organization structure of PTT Post.

Figure 3.1: Organizational structure of PTT Post in 1995



Notes:

Source: adapted from Groot and Smidt (2000).

In 1995, business unit Letters has been reorganized into two different parts, i.e., sorting and distribution, because both functions needed a different focus. The sorting part needed a stronger focus on efficiency, and the distribution part needed a stronger focus on quality and customer service. The distribution part of business unit Letters is organized into 27 geographical areas, which are all responsibility centers. The area manager is responsible for the performance of the entire area. Each area has a number of locations (10 till 30) that collect postal items, sort the postal items on zip code and deliver the products to the customers. Next to the locations, each area has a service center and a customer service center. In the service center supporting, planning, and financial functions are organized. In the customer center contacts with customers are managed. Customers can, for example, ask for faster delivery, and bundled delivery. The 27 area managers are the unit of analyses. All areas have homogeneous tasks and responsibilities. Areas only differ in, for example, size, and geographical location. The processes of the company are labor intensive. Therefore, the critical resources of the areas are employees. Of the total costs measure in the contracts of the managers approximately 85 percent are employees' salaries.

The organization is chosen for a number of reasons. First, in 1996 the organization made an important and substantial change in the PMS of the area managers. Further, quality strategy is important for the organization. Quality strategy is a moderator for the relationship between non-financial performance measures use and financial performance (Ittner and Larcker, 1995, 1997). A third reason is that data requirements needed a substantial number of homogeneous responsibility centers. This homogeneity with respect to, for example, the activities performed, and technology used, is a relative advantage, because less opportunities for local history exist (Mohr, 1995).

The data collection process finished at the end of 1997, since in 1998 the sorting part of the BU started to automate the sorting function, which led to problems that also influenced the distribution function.

3.3 Change in the PMS

From 1988 onwards the company used contracts for different management layers. With these contracts PTT Post started to make managers more accountable, give more financial transparency, and decentralize responsibilities to lower management levels (Groot and Smidt, 2000). Before 1996, area managers were evaluated and rewarded based on costs, revenues, and process quality measures. Higher management started a program to promote quality initiatives throughout the whole organization. The European Foundation of Quality Management (EFQM) model was used to integrate these initiatives. The EFQM-model is a model, comparable with the Balanced Scorecard, in which relationships between different performance dimensions can be examined, and that can help implement a quality strategy. In 1996, the area managers were evaluated for the first time on 5 of the 9 parts of the EFQM-model. These were management of personnel, management of processes, worker satisfaction, customer satisfaction and financial results. Worker satisfaction and customer satisfaction are the new performance dimensions added to the contracts. These two dimensions determine roughly 30 percent of the area managers' evaluation. Table 3.1 gives an overview of the (added) performance dimensions used in the management contracts.

Table 3.1: (Added) performance measures in the contract of the 27 area managers

Performance dimensions in the contract before 1996	Performance dimensions in the contract after 1996
Costs	Costs
Revenues	Revenues
Process quality	Process quality
	Customer satisfaction
	Worker satisfaction

Before the change in the PMS the company had no formal system to continually improve customer and worker satisfaction.

“Sometimes you talked with a customer, or heard complaints, but we never really examined what customers thought of us” (area manager).

Apart from the change in the items of the contracts of the managers no substantial changes took place in the PMS in the period under study. For example, the distribution director who evaluates the area managers is the same person during the period under study. Therefore, the evaluation style is expected to be the same before and after the change.

3.4 Goals of the change in the PMS

As mentioned before, the distribution part of the business unit needed a stronger focus on quality and customer service. This was necessary to prepare the company on the anticipated liberalization of the post market. The EFQM-model was chosen to integrate these initiatives. With this model the managers have an instrument to make tradeoffs between different dimensions of performance. Therefore, this model can be

considered analogue to a Balanced Scorecard. One area manager made the following pertinent remark:

“I see the EFQM-model as a number of saucers that I should keep in the air” (area manager).

There was agreement among the interviewed people that within the organization the most critical factor to achieve improvements in performance is through increased employee satisfaction. The following quote from an internal document called “the power of quality” supports this: “In service, there is a direct relationship between quality and motivated employees”. When employees are more satisfied there will be less absenteeism. As a result, less overtime and less temporary workers will be needed. This will lead to a better process quality, because relatively more mistakes tend to be made in overtime and by temporary workers. Improved process quality, in turn, may lead to a higher customer satisfaction, and lower costs (rework). Better process quality and a higher customer satisfaction may also lead to an increase in revenues. Evaluating managers on worker satisfaction and quality is also congruent with the oft-cited critical components for a successful quality oriented strategy, i.e., employee involvement and process improvement (Ittner and Larcker, 1995).

3.5 Performance measurement cycle

This section describes the performance measurement system of the area managers. The different parts of the system, that is the management contract, the chain contract and the bonus system, are described.

The 27 area managers are evaluated yearly. In the beginning of each year, agreements are made between the director of distribution and each individual area manager about the individual performance measures. At the end of the year each area manager's performance is evaluated. Although area managers are not forced to copy these management contracts to their location managers, service center managers, and customer center managers, all interviewed area managers used similar items to evaluate their subordinates. Additionally, one area manager puts his own targets at the top of the contracts of his subordinates to get an integrated focus. Also, the manager of the customer center is held more responsible for the results of the customer satisfaction survey.

The formal evaluation is subjective, based on reaching the targets. The director of distribution evaluates the area managers. This evaluation mark can be adapted if the area managers can make it clear that there are circumstances why they did not reach the targets. They have to support these circumstances by facts. This suggests that the director of distribution has a 'profit conscious' evaluation style (Hopwood, 1972).

History shows that the categories of the evaluation have a normal distribution in most years. The evaluation mark is the basis for future promotions, salary increases, and for 50 percent of the bonuses. The range of the evaluation mark is from 1 till 5, where 3 means on target.

“Getting a 2 means that you get one or two years to improve. Getting a 1 means that you can look for another job. If you get a 4 this means in general that you can make promotion within the company” (contract manager).

In the management contract there are a number of measures completely under control of the area manager and measures that are partly influenced by the other functions of the business unit Letters, i.e., the sorting and transport function. For these latter measures, from 1994 onwards the area managers make agreements with the sorting managers in a so-called chain contract. When these agreements are made the director of sorting and distribution are also present. In the chain contracts, agreements are made about the % of products that are delivered within a certain time period from the sorting centers to the areas. Although the chain contract is not formally input for the career prospective, salary increase or the bonus, it has become more important throughout the years. The outcomes of this chain contract can also be a factor in the subjective evaluation of the area manager by the distribution director. Of course a good functioning chain is important for the area manager as well.

Half of the manager's bonus is based on the management contract. The other half is based on the performance of the business unit as a whole. The reason that part of the bonus relies on the business unit performance is to motivate cooperation between areas. The maximum amount of the bonus is a substantial part of the base salary.

3.6 Controllability of measures by area manager

Although the respondents of the worker satisfaction survey are the workers in the locations, service centers, and customer centers of each area, by adding the measure to the contract the area manager has an incentive to develop programs to improve worker satisfaction. One example of what an area manager did to improve worker satisfaction, was that employees from the locations could give their top three of elements most important for their satisfaction. After reaching agreement the location managers had to work on these elements and give monthly feedback to the workers. Another manager stressed the importance of selecting location managers who know how to cope with people.

“Although not all his selection test results were very good, the fact that he also was a top-referee in soccer confirmed me of his capabilities” (area manager).

The outcomes of the customer satisfaction survey can be influenced by, for example, after sales service, handling complaints, product innovations, and looking for solutions of problems customers have getting their products in time.

The quality measure is an on-time delivery measure. This measure is completely controllable by area managers.

The cost measure is partly exogenous for the area manager because the number of products delivered, which is not controllable by the area manager, is the largest cost driver. This number of products is taken into account in the target setting process. The area manager can influence the cost measure by initiating efficiency programs. Cost savings from these programs can be used to invest in, for example, customer satisfaction or worker satisfaction.

The revenue item in the contract consists of all parts of revenues that are controllable by the customer service center of the area. Since the headquarters of the business unit arranges contracts with large customers² that are located in different parts of the Netherlands, these are excluded from the area manager's revenue measure. Although for a part of their products the company is a monopolist, they cannot

² Large customers are customers with a yearly revenue of 0.5 million guilders. The marketing and sales department handles these customers.

increase their prices because this is not allowed. An independent agency controls the price setting policy (see section 3.2). Therefore revenues can only be increased by attracting more customers and increasing the number of products per customer.

The fact that large customers are not the responsibility of the areas implies that revenues minus costs is not a proper profit figure in this situation. The area managers are therefore evaluated on separated revenue and cost measures instead of one profit measure.

Chapter 4

Data description

4.1 Introduction

This section describes the data for the empirical analysis. Because both empirical studies in chapter 5 and 6 use the same data set, I define the data and give the descriptives before the analyses.

The research questions addressed in this dissertation are difficult to answer within a cross-sectional research design. First, the incremental information question assesses the impact of lagged non-financial measures on future financial performance. Therefore a lag search is executed, which needs time-series. Second, the evaluation question assesses the impact of adding non-financial measures to the contract of managers. Because non-financial measures are argued to focus on the long-run perspective, a short-run performance measure would not capture the complete performance effects of adding non-financial measures in the PMS (Perera et al., 1997). Therefore a pooled time-series data set of three year monthly observation was collected.

Data about the background of the company (see chapter 3) is gathered through interviews with, the director distribution, the department controller, four area managers, the director information-technology and finance, the contract manager, the manager of quality issues, a location manager, the commercial director, a human resource manager, and a marketing manager.¹ In total, 31 hours of interviews with 21 different managers were held. In addition, internal and external documents, such as contracts, European Foundation of Quality Management documents, were reviewed. Data about the performance measures were either manually or electronically gathered from different departments of the organization.

The first observation is from January 1995 since the 27 areas exist only from that moment. I end the data collection in December 1997 for a number of reasons. First, two areas were put together after 1997. Second, in 1998 there was no worker satisfaction survey and therefore the managers of the areas had less incentive to improve worker satisfaction. Finally, from 1998 onwards the company automated

¹ These interviews were recorded and subsequently transcribed.

parts of the process. This change in technologies might contaminate with effects estimated.

This chapter is structured as follows. Section 4.2 defines the data and the data sources. Section 4.3 discusses advantages and disadvantages of pooled time-series data and explains the fixed effects model. Section 4.4 gives the descriptives of the variables and correlation coefficients between variables.

4.2 Variable definition

The data for the variables used in the study are from the area managers' contracts. These contracts consists of costs, sales, on-time delivery, customer satisfaction and worker satisfaction.² The controller prepares the contracts at the end of the year for the managers' evaluation. For the purpose of this study I need monthly observations and therefore the underlying monthly data that is aggregated to yearly data for the contracts is gathered.

First, the cost variable ($Cost_{it}$) consists of the employee payroll and material and services costs. The measure in the contract is a flexible measure, i.e., it depends on the number of products delivered by the area. The cost measure was collected manually from the contract department. Second, the revenue measure (Rev_{it}) consists of the revenues for which the area managers are held responsible. Large customers that are handled at the headquarters are excluded from the measure. The revenue measure was collected electronically from the financial department. Both $Cost_{it}$ and Rev_{it} are deflated by a consumer price index (CPI_t) to control for inflation. The CPI_t measure is collected from the Dutch Bureau of the Census (Centraal Bureau voor de Statistiek). This measure is defined as an average change in price of goods and services for households. In this context revenues minus costs is not the profit of the area, since the revenue measure in the contract does not include the revenues of large customers. Therefore, revenues and costs are analyzed separately. A theoretical reason to analyze costs and revenues separately is that both might have a different weight in the contract when the measures differ in congruence and noise (Banker and Datar, 1989).

Third, because for worker satisfaction only yearly data are available and the empirical analysis needs monthly observations a proxy for this measure was sought in the literature. Vroom (1964) argued that worker satisfaction is correlated with short-run sickness, whereas the relationship with an overall sickness measure is less clear. Workers who are less satisfied are more frequently sick for a short duration. An overall sickness measure also measures people that are sick for a long time period, whereas an absence frequency measure does not. Therefore, an absence frequency measure was preferred to use as a proxy. In two meta-studies, McShane (1984) and Hackett (1989) found an effect-size of -0.21 between worker satisfaction and absence-frequency. To ascertain the quality of the proxy in the research setting, I correlated the results of the yearly worker satisfaction survey³, i.e., the measure used in the contract with the absence-frequency index, i.e., the proxy used. This correlation was -0.44. This absence-frequency ($Freq_{it}$) measures the number of sickness reports per 100 working days. The measure was manually collected from the human resource department.

² I also tried to find a proxy that is measured each month for the annual customer satisfaction survey. However, candidates like the number of complains, although measured in the company, were not available anymore. Therefore, I cannot use this measure in the analysis.

³ A factor analysis indicated that the yearly worker satisfaction surveys was uni-dimensional.

Finally, a quality measure for the number of rightly supplied products within a certain pre-specified period is available. This measure (OTD_{it}) is a ratio of the number of rightly supplied products within a pre-specified time period divided by the total number of supplied products. Therefore this is an on-time delivery measure. This measure was manually collected from internal reports from the Quality department.⁴

Next to the measures from the area managers' contract, in some models I use control variables. In the models with $Cost_{it}$ as dependent variables, the number of products processed in each period is used as a control variable, because the area manager cannot control this number. This measure Vol_{it} , is defined as the number of products delivered in each time period for each area.

In the model with $Freq_{it}$ as dependent variable, I control for the number of Saturday workers, since McShane (1984) found that age and experience moderated the strength of the relationship between worker satisfaction and absence-frequency. Therefore a control is used for workers on Saturday ($Rsat_{it}$), because these are employees with different characteristics than the regular employees. Most often they are students and are therefore younger and have less experience. Additionally, interviews with human resource managers confirmed the idea that the Saturday workers are less often sick. $Rsat_{it}$ is the number of Saturday workers divided by the total number of workers in each period. In addition, next to using dummies for each month to control for a seasonal pattern in $Freq_{it}$ another control variable used is influenza epidemics (IAZ_{it}). These epidemics are not always in the same month. Using indicator variables, being 1 for months where on average more than 20 out of 10.000 people went to a family doctor in one of four geographical regions and 0 otherwise. This cutoff point of 20 out of 10.000 people is used by the Netherlands institute of primary health care to establish when there is an influenza epidemic.

All variables used in the empirical analyses in chapter 5 and 6, and its definitions are summarized in table 4.1.

⁴ In addition, a second on-time delivery measure was available for a small product group. This variable was also used in the analysis. However, since the variable has hardly any variance the measure was not significant in any model. Therefore I choose not to report results for this variable.

Table 4.1: Variables definitions

Variable	Definition
Rev_{it}	The revenue measure from the contract of area managers of area i in period t . The measure is deflated by CPI_t .
$Cost_{it}$	The cost measure from the contract of area managers of area i in period t . The measure is deflated by CPI_t . The cost measure consists of employee salaries, and material and service costs.
CPI_t	The Consumer price index in period t from the Dutch Bureau of the Census (Centraal Bureau voor de Statistiek).
$Freq_{it}$	The absence-frequency index, measured by the number of sickness reports per 100 working days of area i and period t .
OTD_{it}	The ratio of the number of products supplied at the right place within a pre-specified time period divided by the total number of products of area i in period t
Vol_{it}	The number of products of area i in period t .
$Rsat_{it}$	The ratio of the number of workers on Saturday from the total number of workers of area i in period t .
IAZ_{it}	An indicator variable that is 1 when more than 20 out of 10.000 people went to a family doctor in one of four geographical regions, and 0 otherwise.

4.3 Pooled time-series data

4.3.1 Introduction

Data that consists of multiple measurement of a sample of subjects, for example individuals or organizations, over a period of time is called panel data or pooled time-series data. The distinction between panel data and pooled time-series is that pooled time-series data have fixed intervals between each successive observation and panel data have not (Sayrs, 1989). This section describes pooled time-series data, its statistical problems, and potential solutions.

4.3.2 Properties of pooled time-series data

The use of pooled time-series data has a number of advantages above cross-section or time-series data. The first obvious advantage is that it usually increases the number of observations. This not only increases the power of the tests, but has also some statistical advantages, such as mitigating the problem of multicollinearity, and increased variability (Baltagi, 1995).

A second advantage is that it can control for individual heterogeneity between cross-sections (Baltagi, 1995). That is parameters of independent variables are not constant for each cross-section and time period. For example unobserved cross-sectional characteristics, such as the quality of the management, marketing or manufacturing strategies, can moderate the examined relationship (Huselid and Becker, 1996; Bowen and Wiersema, 1999). Time-series and cross-sectional studies

that do not control for heterogeneity can give biased results.⁵ Another advantage is the extra possibilities available to mitigate the problem of omitted variables in the specification. (Hsiao, 1990). For example in equation (4.1):

$$y_{it} = \alpha + \beta_1 * x_{it} + \beta_2 * z_{it} + \varepsilon_{it} . \quad (4.1)$$

Where:

y_{it} is the dependent variable,

x_{it} is an observable variable,

z_{it} is an unobservable variable and,

i is an subscript for each cross-sections and t for each time period.

When z_{it} correlates with x_{it} then the OLS coefficients are biased. When z_{it} is stable over time, but differs per cross-section, then the first differences⁶, i.e., a change model instead of a level model, of the regression drops the effect of the unobserved variable out, that is $z_{it} - z_{it-1} = 0$. Thus equation (4.2):

$$y_{it} - y_{i,t-1} = \beta_1 * (x_{it} - x_{i,t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1}) . \quad (4.2)$$

gives unbiased and consistent estimates of β_1 .

In addition, when omitted variables are stable for cross sections, but change over time, a regression with deviations from the mean, across cross-sections at a given time (y_t), can give unbiased estimates of β_1 .

$$y_{it} - y_t = \beta_1 * (x_{it} - x_t) + (\varepsilon_{it} - \varepsilon_t) \quad (4.3)$$

Because the research questions examined in this dissertation explore changes over time, and therefore equation (4.3) would remove effects we are interested in, this latter method is not used in this dissertation.

Potential disadvantages of pooled time-series data are the selectivity and heterogeneity bias (Hsiao, 1990). The selectivity bias is the non-random drawing of the sample. Although this is theoretically a problem, the problem is not bigger than for example in survey research in which subjects can choose to respond or not. Also, the research design chosen in this dissertation, i.e., quasi-experimental design is non-random itself.

The process of combining cross-section data with time-series data is called pooling. Before pooling observations, analysis has to make clear whether this is allowed. Pooling data when this is not allowed leads to the heterogeneity bias, i.e., parameters are not the same for all cross-sections or time periods. The fixed effects model, described in the next section mitigates this problem.

⁵ Bowen and Wiersema (1999) argued that heterogeneity in time in cross-sectional research is implicitly addressed when generalizability of the results of the study to other time periods is questioned.

⁶ Using intercept for each cross-section is an alternative method that leads to the same effect (see section 4.3.3).

4.3.3 Fixed effects model

The constant coefficient model, this is the standard ordinary least squares model, assumes that all coefficients are the same for each cross-section in the pool (Says, 1989) and is represented in the following way:

$$y_{it} = \alpha + \beta x_{it} + \varepsilon_{it} \quad (4.4)$$

The drawback of this model is that it ignores variation that is specific to certain cross-sections or time periods. This problem will reveal itself in heteroscedastic error terms. A model that accounts for both cross-sectional and time specific variation is the fixed-effect model (or least square dummy variable) model.⁷ It allows parameters to vary across cross-sections and across time. This can be represented schematically in the following way:

$$y_{it} = \alpha_{it} + \beta_{it} x_{it} + \varepsilon_{it} \quad (4.5)$$

An F-test can formally identify the appropriate model (Baltagi, 1995; Hsiao, 1990). First, test if the parameters are constant for each cross-section (μ), and each time period (λ):

$$H_0: u_i = u_j = u_{n-1} \text{ and } \lambda_1 = \lambda_2 = \dots = \lambda_{t-1}$$

Where:

$$F_0 = [(RSS - USS) / N - 1] / [USS / NT - N - K]$$

has an F-distribution with N-1, N(t-1)-k degrees of freedom.

RSS is the restricted model with a constant intercept and parameters, and USS is the unrestricted model with dummies for each period and each cross-section. If H_0 is rejected similar F-test for constant parameters for each cross-section or for each time period can be assessed.

The choice between these different models is based on the research question examined, the degrees of freedom available and the F-tests described above (Hsiao, 1990). Because in this dissertation the research question is examining a change in time, and therefore allowing parameters to change over time would absorb the effects I am interested in, whereas differences between cross-sections are not of interest, only the test for non-constant intercepts are executed. This means that I only introduce unit heterogeneity in the models (Beck, 2001). For all models reported in the empirical analysis the specification with dummies for each cross-section was the best choice, based on the F-test discussed above.

⁷ An alternative to the fixed effects model is the random effects model. This model however assumes that the source of heterogeneity, e.g., the quality of the management, is uncorrelated with the independent variables (Hsiao, 1990). In addition, the random effect model is preferred over the fixed effects model when the goal of the study is to generalize results to the complete sample (Baltagi, 1995). Since the first assumption is not reasonable in my context, and I do not attempt to generalize to other samples I use the fixed effects models.

4.4 Description of the data

Tables 4.2 and 4.3 give the descriptives, and correlation matrix of all variables. The appendix gives graphs of the two financial and two non-financial measures for each cross-section.

Table 4.2: Descriptive statistics of cost measure, revenues measure, non-financial measures, and control variables

Variable ¹	Sample ²	Mean	Median	Standard deviation	Minimum	Maximum
<i>Cost_{it}</i>	972	56.35	55.00	13.61	26.78	108.24
<i>Rev_{it}</i>	972	30.39	26.66	14.51	10.62	85.01
<i>Freq_{it}</i>	972	1.36	1.30	0.44	0.46	2.74
<i>OTD_{it}</i>	972	90.78	92.20	6.59	55.60	100.00
<i>Vol_{it}</i>	972	17738	17089	5193	7039	41879
<i>Rzat_{it}</i>	972	0.05	0.05	2.46	0.03	0.08

Notes:

¹ Descriptives for the lagged non-financial measures are not incorporated in the descriptives, because for different models the lag is different.

² 3 years of monthly observations of 27 areas are available ($N=3*12*27=972$).

Table 4.3: Correlation matrix of cost measure, revenues measure, non-financial measures, and control variables

Variable ¹	<i>Cost_{it}</i>	<i>Rev_{it}</i>	<i>Freq_{it}</i>	<i>OTD_{it}</i>	<i>Vol_{it}</i>
<i>Cost_{it}</i>	1.00				
<i>Rev_{it}</i>	0.79 **	1.00			
<i>Freq_{it}</i>	0.05	0.22 **	1.00		
<i>OTD_{it}</i>	-0.12 **	-0.15 **	-0.14 **	1.00	
<i>Vol_{it}</i>	0.87 **	0.77 **	0.23 **	-0.01	1.00
<i>Rzat_{it}</i>	-0.23 **	-0.12 **	-0.06	0.10 **	-0.16 **

Notes:

¹ Where **, * means significant at the 0.01, and 0.05, levels (two-tailed test).

The standard deviation of the variables reflects both cross-sectional variance and variance over time. Further, significant correlation coefficients should be interpreted with caution because seasonality, and for example difference in size between the areas, lead to heterogeneity that can causes spurious correlation.

Chapter 5

Relative and incremental information content

5.1 Introduction

This chapter addresses the first research question, i.e., do non-financial performance measures have relative or incremental information content, or both, beyond financial performance measures in predicting future financial performance. As discussed before (see section 1.1) the accounting literature is ambiguous whether non-financial measures have relative or incremental information beyond financial measures for predicting future financial performance. Although a lot of claims are made that non-financial measures are better indicators for future financial performance than financial measure (Kaplan and Norton, 1992; Banker et al., 2000b), the empirical accounting literature often tests the incremental ability of non-financial measures beyond financial measures (Ittner and Larcker, 1998a; Banker et al., 2000b; Nagar and Rajan, 2001). Therefore, in this dissertation I test both the relative and incremental information content of non-financial measures to explore this contention. I use the data described in section 4.2 for the empirical analysis.

This study contributes to the literature in a number of ways. First, it is the first direct test of the relative information content of non-financial performance measures compared to financial measures. Second, I add to the growing literature that studies incremental information content of non-financial performance measures beyond financial performance for predicting future financial performance. In my context I use different non-financial measures, i.e., worker satisfaction and on-time delivery. I also address the non-linearity of the relationship between financial and non-financial measures.

This chapter is structured as follows. Section 5.2 explains the framework used for the empirical analysis. Section 5.3 gives the research methods and specifications used for the analysis. Section 5.4 gives the empirical results. Section 5.5 ends with a discussion and summary.

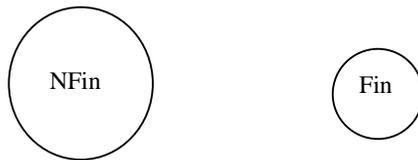
5.2 Framework

5.2.1 Relative information content

Relative information content assesses whether one measure contains more information than another (Biddle et al., 1995). Therefore, relative information content is a stronger assumption than incremental information content, i.e., two measures can have incremental information content beyond each other. Relative information content reflects differences in incremental information content (Biddle et al., 1995)¹. Relative information content is an important question when it is costly to gather information and therefore choices have to be made which measure to supply (Biddle et al., 1995). These costs might not only represent costs in money but also in time spent of managers to use the information. In addition, ranks of different information sources can be provided.

Relative information content can be graphically represented as in figure 5.1.

Figure 5.1: Graphical representation of relative information content



In figure 5.1 the two circles represent the variance explained of future financial performance by financial and non-financial measures. The two measures represent each a part of the future financial performance, but in this case the non-financial measure (NFin) explains more than the financial measures (Fin). Therefore the non-financial measure has a larger information content. The two circles are sketched as non-overlapping, but could equally well overlap.

Relative information content of alternative performance measures is often assessed for valuation purposes, for example, comparing earnings and cashflows to predict future stock returns (Dechow et al., 1994). However, Biddle et al. (1995, p.6) argued that “relative information content comparison [...] could be useful when evaluating alternative performance measures for internal evaluation and control”. For contracting purposes more than one performance measure can be used in the contract. Assessment of relative information content can be an input for assigning weights to different measures for performance evaluation purposes. For example, when agents are risk-neutral or measures are noiseless, principals choose the weights of different performance measures in the contract to maximize congruity (Datar et al., 2001). Since relative information content assesses which measure better predicts future financial performance it can be considered as a comparison of the congruity of different measures.

Kaplan and Norton (1992) argued that non-financial performance measures are better indicators for future financial performance than financial measures. This claim is based on a number of different notions. A non-comprehensive enumeration is that non-financial measures are more actionable, more directly traceable to the strategies of the firm, work well with high-technology manufacturing systems like Automatic

¹ This implies that when measure 1 contains more information than measure 2, per definition measure 1 provides incremental information beyond measure 2.

Manufacturing Technologies and JIT, facilitate continuous improvements (Fisher, 1992), focus attention to causes rather than effects (Singleton-Green, 1993), focus attention to the long-run perspective (Kaplan and Norton, 1992), and give less room for dysfunctional behavior (Fisher, 1992).² Although a lot of anecdotal evidence has been provided to support these claims there is no direct test of the assumption to date.³ Therefore, I compare the relative information content of non-financial measures with financial measures.

Based on the claims in front of non-financial measures given above, I expect the non-financial performance measures have more information content than financial performance measures. In addition, I expect that non-financial measures have higher relative information content compared to financial measures when the number of lags increases. The rationale for this expectation is that non-financial measures are considered to be leading indicators for financial measures (Kaplan and Norton, 1992). Therefore, adding lagged measures in the models to predict future financial performance should favor the non-financial measures. Thus, I assess the relative information content of non-financial and financial measures with different (cumulative) lags.

Table 5.1 summarizes the predictions derived.

Table 5.1: Predictions for the relative information content analysis

Non-financial measures have higher information content than costs for predicting future costs.

Non-financial measures have higher information content than revenues for predicting future revenues.

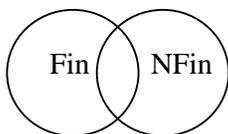
Relative information content of non-financial measures compared with costs increases when (cumulative) lags are added.

Relative information content of non-financial measures compared with revenues increases when (cumulative) lags are added.

5.2.2 Incremental information content

Incremental information content is assessed to indicate whether information is helpful to make better predictions beyond an existing piece of information (Biddle et al., 1995). This can graphically be represented as in figure 5.2.

Figure 5.2: Graphical representation of incremental information content



² Scarce evidence for some of these claims is available from a survey of Van der Stede et al. (2000) of 128 US and Belgian manufacturing managers.

³ Nagar and Rajan, (2001) argued that they compare the relative ability of financial quality measures, e.g., external failure costs, and non-financial quality measures, e.g., defect rates. The tests executed however are incremental tests and do not test directly the relative ability.

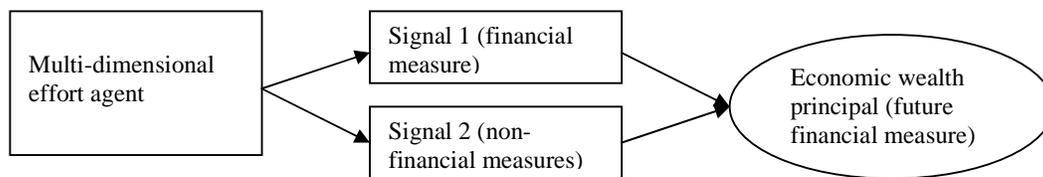
Again, in figure 5.2 the circles represent the variance explained of future financial performance of respective the financial (Fin) and non-financial performance (NFin) measures. The figure indicates that both the financial and non-financial variable explain a part of future financial performance. The non-overlapping part of the non-financial performance is the incremental information content of the non-financial measures beyond the financial measure, and vice versa.

Ittner et al. (1997, p. 233) state that Holmstrom's informativeness principle implies that:

“the bonus contract should place non-zero weight on any performance measure that provides incremental information content about the dimensions of managerial action that the owner wishes to motivate. (initials added)”

In the context of financial and non-financial measures this implies that the non-financial measures should have a non-zero weight in the reward system if it gives information content about the dimensions of managerial action that is not represented in the financial performance measure. This can graphically be represented as in figure 5.3.

Figure 5.3: Graphical representation of informativeness principle in the context of financial and non-financial measures



Managers allocate their effort about multiple dimensions. Because this effort is not directly observable, performance measures are used as signals for this effort. When the non-financial measure, i.e., signal 2, provides incremental information content about the manager's effort beyond the financial measure, i.e., signal 1, the non-financial measures should have a non-zero weight in the reward system.

The concept of incremental information content implies that, to be beneficial the non-financial measure can have either a negative or positive relationship with future financial performance after controlling for the lagged financial measure. Both are helpful to make better predictions. However, for contracting purposes, i.e., to assign the weight to the different measures, it is relevant to know the direction of the relationship. Consequently, this is assessed.

Non-financial performance measures are argued to lead to a long-term orientation of managers (Kaplan and Norton, 1992; Van der Stede et al., 2001). Additionally, non-financial measures better capture the strategic orientation of the organization (Kaplan and Norton, 1996). Therefore, non-financial measures are expected to have additional information beyond lagged financial measures. However, the effects of lagged non-financial performance might be different for either future costs or future revenues. Improving non-financial measures often costs money first because investments must be made. Training employees for example can raise customer satisfaction measures, and improved quality can be reached by redesigning processes. Additionally, improved non-financial measures might also lead to a higher future cost level to maintain the higher level of the non-financial measure. In contrast, higher non-financial measures might lead to lower future cost levels when for example better

quality measures lead to lower future costs, due to less scrap, rework on faulty products. Earlier research is also inconclusive about the direction of the incremental ability of non-financial measures to predict future costs. Banker et al. (2000b) find that one of their customer satisfaction measures has a negative impact on future costs. Ittner and Larcker (1998a) however, find no relationship between the customer satisfaction measure and future expenses. Therefore lagged non-financial measures can have a differential effect on future costs, and no expectation is formulated for the incremental information content of non-financial measures beyond lagged financial measures.

As argued above non-financial performance measures are claimed to lead to a long-term orientation of managers and better capture the strategic orientation of the organization. All available literature suggests that non-financial measures have a positive impact on future financial performance after controlling for current financial measures (Ittner and Larcker, 1998a; Banker et al., 2000b; Nagar and Rajan, 2001). Therefore, non-financial measures expected to have additional positive information to lagged revenues measures.

The expectations derived are summarized in table 5.2.

Table 5.2: Expectations for incremental information content

Non-financial measure → financial measure	Expectation ¹
Worker satisfaction → future costs	?
On-time delivery → future costs	?
Worker satisfaction → future revenues	+
On-time delivery → future revenues	+

¹A + means that the non-financial measure is expecting to have a positive impact on future financial performance after controlling for the lagged financial performance. A ? means that the non-financial measure is expecting to have a differential impact on future financial performance after controlling for the lagged financial performance.

5.3 Specification

5.3.1 Relative information content

Relative information content is assessed through a non-nested model procedure. It is not possible to test non-nested models with an F-test. Therefore, Vuong (1989) designed a likelihood statistic that compares two non-nested models with each other. He shows that this likelihood statistic has a normal distribution. The test is able to assess which of the competing models has more information content, i.e., is closer to the “true” model. Intuitively, the Vuong (1989) statistic compares the R^2 ‘s of the two competing models, and assesses whether the difference is significant.

The z-value of the likelihood statistic can easily be computed by the following procedure (Dechow, 1989, p. 39). First, execute the following regressions:

$$F_t = \alpha + \beta * F_{t-1} + \varepsilon \quad (5.1)$$

and,

$$F_t = \alpha + \beta * NF_{t-1} + \varepsilon_t. \quad (5.2)$$

Where:

F_t = a financial performance measure in period t ,
 F_{t-1} = a lagged financial performance measure, and
 NF_{t-1} = a lagged non-financial measure.

Then compute m_{it} for each observation through $m_{it} = 0.5 \log^* \left(\frac{RSS_f}{RSS_{nf}} \right) + n/2 * \left[\frac{(\varepsilon_f)^2}{RSS_f} - \frac{(\varepsilon_{nf})^2}{RSS_{nf}} \right]$, where RSS_f is the residual sum of square, and ε_f are the residuals of (5.1), and RSS_{nf} is the residual sum of square, and ε_{nf} are the residuals of (5.2). The sum of m_{it} is the Vuong statistic.

The next step is to show whether this statistic is significant, that is to compute the standard deviation from the statistic. This can be found through regressing m_{it} on unity. The t-statistic from this coefficient times $((n-1)/n)^{1/2}$ gives the z-value. Since it uses the lagged financial performance measures as the benchmark a positive z-value indicates that the non-financial performance measures has more information content. For details of the test see Vuong (1989) and Dechow (1994).

For the revenues, the two models that are assessed against each other are:

$$Rev_{it} = \lambda_{ri} + \eta_{r1} * Rev_{it-l} + \sum_{k=1}^{11} \eta_{r2k} * MONTH_k + \varepsilon_{rit} \quad (5.3)$$

and,

$$Rev_{it} = \lambda_{ri} + \eta_{r1} * OTD_{it-l} + \eta_{r2} * Freq_{it-l} + \sum_{k=1}^{11} \eta_{r3k} * MONTH_k + \varepsilon_{rit}. \quad (5.4)$$

Where:

Rev_{it} = the revenue measure in the contract of area managers, that consist of all revenues that areas can influence. The measure is deflated by CPI_t ,
 CPI_t = Consumer price index from the Centraal Bureau voor de Statistiek (the Dutch Bureau of the Census),
 $Freq_{it-l}$ = the frequency index with up to 12 lags,
 OTD_{it-l} = the on-time delivery measure with up to 12 lags,
 $MONTH_k$ = an indicator being 1 for month k and 0 otherwise,
 t = index of period,
 i = index for each area,
 l = number of lags.

In both models dummy variables are used for each area. These so-called fixed effects models are used to control for heterogeneity due to variables that are unique for each area, and are constant in time (Hsiao, 1990). Obvious examples in this context are the size of the area, leadership of the area manager, and socio-economic factors, that differ between the areas. Using dummies for eleven months controls for a seasonal pattern.

To assess the relative information content of the cost measure with the non-financial measure the following models are compared:

$$Cost_{it} = \lambda_{ci} + \eta_{c1} * Cost_{it-1} + \sum_{k=1}^{11} \eta_{c2k} * MONTH_k + \eta_{c3} * Vol_{it} + \varepsilon_{cit} \quad (5.5)$$

and,

$$Cost_{it} = \lambda_{ci} + \eta_{c1} * OTD_{it-1} + \eta_{c2} * Freq_{it-1} + \sum_{k=1}^{11} \eta_{c3k} * MONTH_k + \eta_{c4} * Vol_{it} + \varepsilon_{cit}. \quad (5.6)$$

Where:

$Cost_{it}$ = the cost measure from the contract of area managers, consists of employee costs and cost for material and services. The measure is deflated by CPI_t ,

Vol_{it} = the number of products of area i in period t .

The cost measure in the contract is a flexible measure. It can vary by the number of products that are handled in the area in each period. Therefore, in the cost model a control for the number of products handled in each period (Vol_{it}) is used. Again, dummy variables are used for each area, and for eleven months.

Since I expect that relative information content of non-financial measures increases when more lags are used in the model, the models are compared using different cumulative lags. Thus, I assess the models with t-1, t-1 and t-2, etc. I stop adding lags to the models at 12 lags because at that point only 10 observations per parameter are available.

5.3.2 Incremental information content

Incremental information content is assessed by observing the t-statistic of the parameter of the lagged non-financial measure (β_2) in the following regression:

$$F_t = \alpha + \beta_1 * F_{t-1} + \beta_2 * NF_{t-1} + \varepsilon. \quad (5.7)$$

Where:

NF_{t-l} = a lagged non-financial measure, with l lags,
and all other variables are as before.

The lagged dependent variable is used as an explanatory variable in the specification to ascertain whether the non-financial measures give additional information over the financial measure from the period before.

Ittner and Larcker (1998a) and Banker et al. (2000b) discuss the proper specification to assess the incremental information content of financial and non-financial performance measures. Because both level models and change models have pros and cons (see section 5.3.3) they estimate both specifications to explore the relationship. This strategy is followed in this chapter as well.

The assumption that non-financial performance measures are indicators for future financial performance is a central issue in the Balanced Scorecard concept (Kaplan and Norton, 1992). However, non-financial measures do not need to have an immediate impact on financial measures, but might have an impact on them only a few periods later. Unfortunately, no formal theory is available that can indicate the length of this so-called "lag". Therefore, a specification search is executed to explore this lag from the data. This specification search is executed for the first time in

management accounting research in Banker et al. (2000b), but is much more familiar in economic research.

The lag search is based on the following equation, an unrestricted finite distributed lag model (Greene, 2000):

$$y_t = \alpha + \sum_{i=0}^q \beta_i X_{t-i} + \varepsilon_t \quad (5.8)$$

In which y_t is the dependent variable in period t , in this context the financial measure, and the second term of the right hand site of (5.8) are the lagged independent variables, in this context the lagged non-financial measures, up to lag i .

The specification search is based on the Akaike information criterion (AIC) (Akaike, 1973). The AIC is comparable with the adjusted R^2 in that it penalizes losses in degrees of freedom and rewards increased fit. The AIC criterion differs from the adjusted R^2 in that it penalizes losses in degrees of freedom more heavily (Greene, 2000). A natural alternative for the AIC criteria would be a number of sequential F-tests until the total number of lag coefficients is 0. This would however inflate the significant level (Greene, 2000). The equation used to compute the AIC is:

$$AIC = -2 * l / n + 2 * k / n \quad (5.9)$$

In which k is the number of estimated parameters, l is the value of the log likelihood and, n is the number of observations.

The lag specification search chooses the number of lags where the AIC is minimal. Due to problems with degrees of freedom a maximum number of lags must be pre-specified over which the search is executed (Winker, 2000). This is 12 months in this analysis.

The coefficients of the different lagged non-financial measures are not interpretable due to high multicollinearity. Therefore after finding the optimal number of lags in the model, a Principal Component (PC) analysis reduces the lagged non-financial measures to one construct (Greene, 2000). The coefficient of this construct, i.e., lagged non-financial performance, is the long-run average impact of the non-financial measure on contemporary performance.

Thus, the following equation is estimated to explore the relationship between $Cost_{it}$ and the lagged non-financial performance measures:

$$Cost_{it} = \lambda_{ci} + \eta_{c1} * Cost_{i,t-1} + \eta_{c2} * AvFreq_{i,t} + \eta_{c3} * AvOTD_{i,t} + \sum_{k=1}^{11} \eta_{c4k} * MONTH_k + \eta_{c5} * Vol_{it} + \varepsilon_{it} \quad (5.10)$$

Where:

$AvFreq_{i,t}$ = averaged lagged performance of $Freq_{it}$. Estimated by averaging the l lagged measures,

$AvOTD_{i,t}$ = averaged lagged performance of OTD_{it} . Estimated by averaging the l lagged measures,

and other variables are as before.

Indicator variables are used for each area, i.e., the fixed effects model and for eleven months to control for a seasonal pattern.

The following equation is estimated to explore the relationship between revenues and the lagged non-financial performance measures:

$$\text{Re } v_{it} = \lambda_{ri} + \eta_{r1} * \text{Re } v_{i,t-1} + \eta_{r2} * \text{AvFreq}_{i,t} + \eta_{r3} * \text{AvOTD}_{i,t} + \sum_{k=1}^{11} \eta_{r4k} * \text{MONTH}_k + \varepsilon_{rit}. \quad (5.11)$$

Where:

All variables are as before.

5.3.3 Statistical considerations

The Jarque-Bera statistic does not reveal large problems with non-normality of the residuals. For example, in one model the Jarque-Bera statistic was significant at the 5% level for only 9 of the 27 areas. Because of these results and in combination with the large sample size I choose not to transform the data for non-normality.

If the disturbances of each equation are uncorrelated there is no relationship between the equations and OLS is appropriate (Pindyck and Rubinfeld, 1998). When this is not the case the parameters are unbiased and consistent, but inefficient, and the estimates of their variance could be biased (Parks, 1967). Correlation between different cross-sections can be expected, since all areas are from one company. Therefore, a similar reaction of an external shock can lead to correlated residuals between the cross-sections. Therefore, the results reported are all from the seemingly unrelated regression (SUR)-model.^{4,5} In addition, next to the problem of correlated residuals the SUR-model also mitigates the problem of cross-sectional heteroscedasticity. Heteroskedasticity can be expected because larger areas may have larger variance.

One of the most problematic specification issues is the choice for level or change models. Ittner and Larcker (1998a), and Banker et al. (2000b) discuss the proper specification to assess the incremental information content of non-financial measures. The change model has the advantage that it controls for omitted variables that are stable over time. However, it also assumes a constant relation between the change in the different lags and the dependent variable (Banker et al., 2000b), i.e., it expects that the impact of the non-financial measure at t-2, and t-1 have the same impact on the future financial performance measure. This is obviously not necessary (Banker et al., 2000b). Additional, problems with change variables are high unreliability⁶ and correlation with its components⁷ (Cohen and Cohen, 1983). In addition, in my opinion most omitted variables are constant in time and therefore are picked up by the fixed

⁴ Efficiency gains of SUR are higher when correlation between residuals increase. In contrast, the efficiency gains are low when in all models the same independent variables are used, data is trended and there is greater correlation among the independent variables (Doran and Griffiths, 1983). In this case, for each area the same independent and control variables are used, data of the control variables is trended and correlation between independent variables is low.

⁵ Correlations between the residuals of the different cross-section range between -0.6 and 0.8, indicating that using OLS might be seriously flawed.

⁶ The unreliability of a change score decreases when the correlation (ρ) between its component increases. Let assume that the both components (Y_1 and Y_2) have the same reliability. Then the reliability of the change score ($Y_2 - Y_1$) is $(\text{reliability}(y_1, y_2) - \rho(y_1, y_2)) / (1 - \rho(y_1, y_2))$ (Allison, 1990). Thus, when the reliability of the level variables is 0.7 and their correlation is 0.3. The reliability of the change variable is 0.57.

⁷ This is the "regression to the mean" effect. High values of a variable will probably go down in the next period.

effects model (see section 4.3.3). Therefore, although both level and change models are reported I rely on the level models for the overall conclusions.^{8,9}

I re-estimated all models without observations that had residuals that are higher than 3 standard deviation from the mean (Hair et al., 1998). The maximum number of outliers for a model was 10. These outliers do not influence the results reported. T-statistics are in general higher for these regressions without outliers.

Although in sections 5.2.1 and 5.2.2 I derive expectations for some tests, these are highly exploratory. Therefore, for reasons of conservatism I report two sided significance tests.

The R^2 -adjusted are high in all models. These high R^2 -adjusted are mainly driven by the fixed effects for each area that control for omitted variables that are constant in time, such as size of the area, and indicator variables for the months to control for seasonality.

5.4 Empirical results

5.4.1 Introduction

The descriptives of the variables used in this section are given in table 4.2 and 4.3. Section 5.4.2 gives the results for the relative information content analysis. Next, Section 5.4.3 gives the results from the lag search procedure. Section 5.4.4 reports the incremental information content analysis. Section 5.4.5 discusses the managerial significance of the incremental information content of the analysis. Finally, section 5.4.6 analysis whether incremental information content of non-financial measures is non-linear.

5.4.2 Relative information content

The results of the relative information content analysis are reported in table 5.5.

⁸ Interestingly, Lambert (1998) criticizes the use of level models in the discussion of Ittner and Larcker's (1998a) paper. He only gives advantages of change models, i.e., mitigating omitted variables, without discussing the disadvantages. Nagar and Rajan (2001) also favor change models because it mitigates the problem of omitted variables.

⁹ A final difference between level and change models is that the results for these two different models do not have the same meaning. Therefore they are not additive support for the same hypothesis.

Table 5.5: Relative information content of non-financial performance measures (level models)

Number of observations ¹	<i>Cost_{it}</i> -model z-value ^{2,3,4} (p-value ⁵)	<i>Rev_{it}</i> -model z-value ^{2,3,4} (p-value ⁵)
1 (945)	-1.97 (<0.03)	1.72 (n.s.)
2 (918)	-1.98 (<0.03)	0.59 (n.s.)
3 (891)	-1.66 (n.s.)	0.46 (n.s.)
4 (864)	-1.15 (n.s.)	-0.09 (n.s.)
5 (837)	-1.06 (n.s.)	0.74 (n.s.)
6 (810)	-0.83 (n.s.)	0.02 (n.s.)
7 (783)	-0.09 (n.s.)	0.08 (n.s.)
8 (756)	-0.55 (n.s.)	-0.56 (n.s.)
9 (729)	-0.64 (n.s.)	-0.60 (n.s.)
10 (702)	-0.84 (n.s.)	-0.50 (n.s.)
11 (675)	-0.57 (n.s.)	0.06 (n.s.)
12 (648)	-1.99 (<0.03)	0.28 (n.s.)

Notes:

¹ Since I use a cumulative number of lags in each analysis, the number of observations decreases with 27 observations with each test.

² Z-values are from the Vuong (1989) likelihood statistic.

³ The non-financial measure is used as a benchmark, therefore a + means that the non-financial measure has more information content, and a - means that the financial measures has more information content.

⁴ Results are from the SUR model. However to be able to estimate a SUR model more time periods than cross-sections are necessary. Therefore from the 9th lag onwards results are from White's (1980) procedure.

⁵ Where n.s. means not significant.

The results for the *Cost_{it}*-model show no support for the expectation that non-financial measures are better indicators for future cost than lagged costs. In contrast, in most tests the lagged costs have higher information content than the non-financial measures, although only 3 are significant. In other cases, the models are not distinguishable from each other at a 5% level. Although z-values initially decrease when more lags are used in the test, this pattern is not stable over all tests. Therefore, there is also no support for the second expectation that the information content of non-financial measures increases, compared to the financial measure, when more lags are added.

Also, for the *Rev_{it}*-model, the results are not supporting the claim that non-financial measures have more information content than financial measures. Although in most cases z-values are positive none of them is significant. In addition, there is no indication that the information content of non-financial measures increases when more lags are added.

To ascertain that the significant results for the cost model are not due to an artificial relationship, i.e., autocorrelation between t and t-1, t-2 and t-12, I also report the results for change variables. The results for these models are reported in table 5.6.

Table 5.6: Relative information content of non-financial performance measures (change models)

Number of observations ¹	$\Delta Cost_{it}$ -model z-value ^{2,3,4} (p-value)	ΔRev_{it} -model z-value ^{2,3,4} (p-value)
1 (918)	-3.38 (<0.01)	-4.03 (<0.01)
2 (891)	-4.59 (<0.01)	-4.36 (<0.01)
3 (864)	-4.00 (<0.01)	-4.21 (<0.01)
4 (837)	-3.43 (<0.01)	-4.00 (<0.01)
5 (810)	-3.42 (<0.01)	-4.08 (<0.01)
6 (783)	-3.38 (<0.01)	-3.78 (<0.01)
7 (756)	-3.70 (<0.01)	-4.39 (<0.01)
8 (729)	-3.83 (<0.01)	-4.13 (<0.01)
9 (702)	-3.73 (<0.01)	-4.11 (<0.01)
10 (675)	-4.29 (<0.01)	-4.14 (<0.01)
11 (648)	-6.66 (<0.01)	-4.06 (<0.01)
12 (621)	-5.77 (<0.01)	-3.92 (<0.01)

Notes:

¹ Since I use a cumulative number of lags in each analysis, the number of observations decreases with 27 observations in each step.

² Z-values are from the Vuong (1989) likelihood statistic.

³ The non-financial measure is used as a benchmark, therefore a + means that the non-financial measure has more information content, and a - means that the financial measures has more information content.

⁴ Results are from the SUR model. However to be able to estimate a SUR model more time periods than cross-sections are necessary. Therefore from the 8th lag onwards results are from White's (1980) procedure.

For the change models, in all models with any number of lags the financial measure has more information content than the non-financial model. This suggests that the three significant results in favor of the financial measure in the level models are not due to an artificial relation between t and $t-1$ or $t-2$.

In sum, I interpret the results as no indication that the non-financial performance measures in this context have more information content than lagged financial performance measures. There is also no evidence that non-financial measures have more relative information content compared with lagged financial measurers when the number of lags incorporated in the models increases.

5.4.3 Lag search

In Table 5.7 the results for the lag search procedure for the $Cost_{it}$ and Rev_{it} -model are summarized.

Table 5.7: lag search procedure

Variable ¹	$Cost_{it}$	Rev_{it}
$AvFreq_{it}$ ²	4	1
% variance explained by first Principal Component	57	-
$AvOTD_{it}$ ²	3	5
% variance explained by first Principal Component	61	50

Notes:

¹ The results are from the lag specification search. The Akaike information criterion is used to find the proper specification. The lagged dependent variable (t-1) was used in the model.

² The Principal Component analysis summarizes the number of lagged periods to one construct, i.e., lagged non-financial performance.

The lag search led to the result that in the $Cost_{it}$ -model the best model has a four period lag of $AvFreq_{it}$ and a three period lag of $AvOTD_{it}$. In a Principal Component analysis the first component of $AvFreq_{it}$ explained 57 percent of the variance. For $AvOTD_{it}$ this was 61 percent.

For the Rev_{it} -model, the best model had a one period lag of $AvFreq_{it}$, and a five period lag of $AvOTD_{it}$. The first component of $AvOTD_{it}$ explained 50 percent of the variance.

These results show that non-financial measures can have different lags for the different measures. This result is in contrast with the study of Banker et al. (2000b), which shows a lag of six months for both non-financial measures for cost, revenue, and profit measures.

It is difficult to explain the drivers of the different lags for different non-financial and financial measures. One line of speculation could be that costs are more controllable than revenues, because they are made within each area, whereas revenues are made through customers outside of the areas. This would suggest that lags for the cost measure are shorter than for the revenues. However this conjecture is not supported by the data.

5.4.4 Incremental information content

Table 5.8 reports the results of the analyses of the incremental information content of the non-financial measures beyond the financial measures for predicting future financial performance for the level models.

Table 5.8: Incremental information content of non-financial performance measures (level models)

Variable ¹	Expected Sign	$Cost_{it}$ (n=864 ²)	Expected Sign	Rev_{it} (n=837 ³)
		Coefficient ^{4,5}		Coefficient
$Cost_{t-1}$		0.224 ** (8.62)		
Rev_{t-1}				-0.105 ** (-3.66)
$AvFreq_{it}$?	-1.489 ** (-8.20)	-	-0.642 ** (-11.75)
$AvOTD_{it}$?	-0.085 ** (-13.93)	+	0.085 ** (24.29)
Vol_{it}	+	0.000 ** (17.90)		
R^2 -adj.		0.966		0.984

Notes:

¹ The month dummies and intercepts for each area are considered to be nuisance variables, therefore they are not reported.

² Because 4 lags are the maximum in this model, $(972-(4*27))=864$ observations are available.

³ Because 6 lags are the maximum in this model, $(972-(5*27))=837$ observations are available.

⁴ Where **, * means significant at the 0.01, and 0.05 levels (two-tailed test).

⁵ t-statistics between parentheses. All results are from the seemingly unrelated regression.

Parameters η_{c2} and η_{c3} of equation (5.10) test the incremental information content of non-financial measures beyond lagged costs. The analysis supports the expectation that the non-financial measures have incremental information content beyond the lagged costs. The significant parameter η_{c3} , -1.489 (t=-8.20, p<0.01), suggest that lagged worker satisfaction, measured by $AvFreq_{it}$, has a negative impact on future cost.¹⁰ One explanation for this result is that investments needed to improve worker satisfaction are larger than the returns in the long-run. Further, the lagged on-time delivery measure, $AvOTD_{it}$, has a positive impact on long-run cost, represented by the significant parameter η_{c3} of -0.085 (t=-13.93, p<0.01). One possible explanation is that better product quality needs less rework in future periods and therefore costs decrease. Current Cost, $Cost_{it-1}$ also has a significant positive impact on future costs, 0.224 (t=8.62, p<0.01). This result is in resemblance with Banker et al. (2000b) that also find a positive lagged cost parameter.

Parameters η_{r2} , and η_{r3} of equation (5.11) test the incremental information content of non-financial measures beyond lagged revenues. The analysis supports the expectation that the non-financial measures have incremental information content beyond the lagged revenues. The impact of worker satisfaction, measured by $AvFreq_{it}$, is positive, -0.642 (t=-11.75, p<0.01). That is higher lagged worker satisfaction leads to higher future revenues. This suggests that happy employees will lead to additional revenues in the long-run. Lagged performance on the on-time delivery measure $AvOTD_{it}$, represented by a significant coefficient of 0.085 (t=24.29, p<0.01), is also positive on future revenues. Current revenue is also a significant predictor for future revenues, -0.105 (t=3.66 p<0.01). This result is in contrast with the study of Banker et al. (2000b) that finds a positive impact of lagged revenues on future revenues.

¹⁰ Remember that the proxy is inversely related with worker satisfaction.

In the analysis above the parameters for the lagged non-financial variables are considered to be homogeneous for each cross-section. Rerunning the regression with specific parameters for each cross-section for the $Cost_{it}$ -model, i.e, parameters η_{c2} and η_{c3} in equation (5.10) are cross-section specific, learns that for $AvFreq_{it}$ for 16 areas the parameter is negative and significant, and for 11 it is not significant at the 5% level. For the $AvOGS_{it}$ 22 areas have negative significant parameters, 3 are insignificant, and 2 are significant positive. For the Rev_{it} -model, rerunning the regression with specific parameters for each cross-section, i.e, parameters η_{r2} and η_{r3} in equation (5.11) are cross-section specific, learns that for $AvFreq_{it}$ for 15 areas the parameter is negative and significant, and for 12 it is not significant at the 5% level. For the $AvOGS_{it}$ 17 areas have positive significant parameters, 9 are insignificant, and 1 is significant and negative. These analysis, reported in Table 5.9, suggest that there might consist considerable cross-sectional differences in the incremental information content of non-financial measures even in a relative homogeneous sample of different cross-sections from one organization.

Table 5.9: Cross-sectional differences of incremental information content of non-financial measures

	+	n.s.	-
<i>Cost_{it}-model</i>			
<i>AvFreq_{it}</i>	0	11	16
<i>AvOGS_{it}</i>	2	3	22
<i>Rev_{it}-model</i>			
<i>AvFreq_{it}</i>	0	12	15
<i>AvOGS_{it}</i>	17	9	1

Notes:

This table reports the number of cross-sections that have a positive (+), not significant (n.s.), and negative (-) incremental information content of the non-financial measure.

In sum, the data supports the expectation that the lagged non-financial measures have incremental information content beyond the lagged financial measures. Analyses for the level models suggest that higher lagged worker satisfaction has a negative impact on future cost, i.e., cost increase, and a positive impact on future revenues. Higher lagged quality, $AvOTD_{it}$, leads to decreased future cost and increased revenues. The results for incremental information content of lagged on-time delivery for future revenues corroborates the finding of Nagar and Rajan (2001). However, they do not assess the impact on future costs.

Table 5.10 reports the results of the change models.

Table 5.10: Incremental information content of non-financial measures (change model)

Variable ¹	$\Delta Cost_{it}$ (n=837 ²)		ΔRev_{it} (n=810 ³)	
	Coefficient ^{4,5}		Coefficient	
$\Delta Cost_{t-1}$	-0.372	**		
	(-20.54)			
ΔRev_{t-1}			-0.610	**
			(-29.59)	
$\Delta AvFreq_{it}$	-3.288	**	-0.017	
	(-14.74)		(-0.39)	
$\Delta AvOTD_{it}$	-0.041	**	-0.012	**
	(-6.56)		(-8.90)	
ΔVol_{it}	0.000	**		
	(34.23)			
R^2 -adj.	0.619		0.374	

Notes:

¹ The month dummies and intercepts for each area are considered to be nuisance variables, therefore they are not reported.

² Because 4 lags are the maximum in this model, and a change model is used $(945-(4*27))=837$ observations are available.

³ Because 5 lags are the maximum in this model, and a change model is used $(945-(5*27))=810$ observations are available.

⁴ Where **, * means significant at the 0.01, and 0.05 levels (two-tailed test).

⁵ t-statistics between parentheses. All results are from the seemingly unrelated regression.

For $\Delta Cost_{it}$ the analysis leads to the same results as for the level model. Thus, lagged changes in worker satisfaction leads to larger increases in costs. Improvements in the on-time delivery measure OTD_{it} however leads to decreases in future costs.

Results for the ΔRev_{it} -model are mixed. The impact of a change in the lagged worker satisfaction ($\Delta AvFreq_{it}$) does not lead to a significant change in future revenues. The result for the impact of a change in lagged quality is even more puzzling. A lagged increase in $\Delta AvOTD_{it}$ leads to a decrease in ΔRev_{it} . This implies that customers do not reward the company for improvement in quality with more future revenues.

The variables for lagged non-financial performance, $AvFreq_{it}$ and $AvOTD_{it}$, are measured by averaging the individual lagged measures. However this assumes that the impact of each lagged measures on the current financial measure is equal for each period. This also means that it is not necessarily congruent with the concepts of leading and lagging indicators. The lag search assumes that the impact of the non-financial measure at period $t-1$ is immediately felt in the next period, however it is also possible, even more likely, that the measure at period $t-1$ influences the financial measure not before, for example, period $t+2$. This assumption is not only unlikely but also unnecessary. Therefore, an alternative specification is used in which the lagged non-financial measure is measured by using the factor scores from the Principal Component analysis. These factor scores allow that different lags of non-financial measures can have a different impact on future financial performance. Results for this analysis for the level and change models are reported in table 5.11 and 5.12.

Table 5.11: Incremental information content of non-financial performance measures (level models with factor scores)

Variable ¹	<i>Cost</i> _{it} (n=864 ²)		<i>Rev</i> _{it} (n=837 ³)	
	Coefficient ^{4,5}		Coefficient	
<i>Cost</i> _{t-1}	0.226	**		
	(8.68)			
<i>Rev</i> _{t-1}			-0.081	**
			(-2.82)	
<i>AvFreq</i> _{it}	-1.70	**	-0.715	**
	(-7.05)		(-12.32)	
<i>AvOTD</i> _{it}	-0.101	**	0.127	**
	(13.48)		(22.14)	
<i>Vol</i> _{it}	0.000	**		
	(18.04)			
R²-adj.	0.966		0.984	

¹ The month dummies and intercepts for each area are considered to be nuisance variables, therefore they are not reported.

² Because 4 lags are the maximum in this model, (972-(4*27))=864 observations are available.

³ Because 5 lags are the maximum in this model, (972-(5*27))=837 observations are available.

⁴ Where **, * means significant at the 0.01, and 0.05 levels (two-tailed test).

⁵ t-statistics between parentheses. All results are from the seemingly unrelated regression.

Table 5.12: Incremental information content of non-financial performance measures (change models with factor scores)

Variable ¹	Δ <i>Cost</i> _{it} (n=837 ²)		Δ <i>Rev</i> _{it} (n=810 ³)	
	Coefficient ^{4,5}		Coefficient	
Δ <i>Cost</i> _{t-1}	-0.372	**		
	(-20.39)			
Δ <i>Rev</i> _{t-1}			-0.606	**
			(-28.42)	
Δ <i>AvFreq</i> _{it}	-3.692	**	-0.003	
	(-10.67)		(-0.07)	
Δ <i>AvOTD</i> _{it}	-0.054	**	-0.017	**
	(-6.68)		(-7.58)	
Δ <i>Vol</i> _{it}	0.000	**		
	(35.84)			
R²-adj.	0.618		0.376	

¹ The month dummies and intercepts for each area are considered to be nuisance variables, therefore they are not reported.

² Because 4 lags are the maximum in this model, and a change model is used (945-(4*27))=837 observations are available.

³ Because 5 lags are the maximum in this model, and a change model is used (945-(5*27))=810 observations are available.

⁴ Where **, * means significant at the 0.01, and 0.05 levels (two-tailed test).

⁵ t-statistics between parentheses. All results are from the seemingly unrelated regression.

The results where the lagged non-financial measures are computed with factor scores are qualitatively the same compared to the results with averaging. Although the parameters for the lagged non-financial measures are somewhat higher they are in the same direction and have the same significance level. This can be explained by the fact that in this context the factor loadings are almost the same for each individual lagged measure. This set of results also mitigates the problem that Banker et al. (2000b) give with change models. They argued that the change models assume that each lagged non-financial measure has the same impact on the future financial measure and that this is an unrealistic assumption. In my setting this does not seem to be a problem.

In general, the data indicate that the non-financial measures have incremental information content beyond the lagged financial measure for future financial performance. Worker satisfaction has a negative impact on future costs but a positive influence on revenues. The on-time delivery measure had a positive impact on both costs and revenues.

5.4.5 Managerial significance

To be able to compare the relative impact of the two non-financial measures with the financial measure I report standardized coefficient. In addition, to estimate whether the results found are material I estimate the managerial significance of the parameters of the lagged non-financial measures. These analyses are reported in table 5.13.

Table 5.13: Standardized coefficient and managerial significance of non-financial performance measures

	Standardized coefficient	Managerial significance (%) ¹
<i>Cost_{it}</i> -model		
<i>Cost_{i,t-1}</i>	0.222	
<i>AvFreq_{i,t-4}</i>	-0.037	-0.36
<i>AvOTD_{i,t-3}</i>	-0.031	-1.36
<i>Rev_{it}</i> -model		
<i>Rev_{i,t-1}</i>	-0.105	
<i>AvFreq_{i,t-1}</i>	-0.022	-0.28
<i>AvOTD_{i,t-5}</i>	0.025	2.54

¹ This measure is computed as followed: ((10% of average of independent variable*coefficient) / average dependent variable)*100%.

To be able to compare the amount of incremental information content of the non-financial measures beyond the financial measure, I report the standardized coefficients of the non-financial measures and the lagged financial measures (Hair et al., 1998). This analysis indicates that the incremental information content of the two non-financial measures beyond the financial measure is roughly equal in absolute terms. The incremental information content of the non-financial measures is roughly one-seventh, -0.037 and -0.031 versus 0.224, of the information content of the lagged cost

measure. For the revenue model, the non-financial measures have roughly one-fifth, -0.022 and 0.025 versus -0.105, of the information content of lagged revenues.¹¹

The managerial significance of the incremental information content of the non-financial measures is assessed by the percentage increase of the dependent variable with a 10% increase in the independent variable (column 3), assuming the rest of the variables are constant. This analysis shows that for the cost model, the future costs are 0.36% lower when $AvFreq_{i,t-4}$ is 10% higher, and the future costs are 1.36% lower when $AvOGS_{i,t-3}$ is 10% higher. For the revenue model the future revenues are 0.28% lower when the $AvFreq_{i,t-1}$ is 10% higher, and the future revenues are 2.54% higher when the $AvOGS_{i,t-6}$ is 10% higher. Considering the fact that these measures are ratio measures, and therefore it is almost impossible to increase the non-financial measures with an additional 10% these effects are not very high.

5.4.6 Non-linearity in incremental information content

Ittner and Larcker (1998) find non-linear relationships between customer satisfaction measures and future financial performance. One explanation for these results is that it becomes increasingly more difficult to improve non-financial measures when they are already on a high level. In addition, customers might not value further improvements in, for example, quality when quality is already high. Therefore, I test for non-linearities through an interaction between the change in lagged non-financial measures and the level of lagged non-financial measures.¹² That is the following regressions are estimated:

$$\begin{aligned} \Delta Cost_{it} = & \lambda_{ci} + \eta_{c1} * \Delta Cost_{i,t-1} + \eta_{c2} * AvFreq_{i,t} + \eta_{c3} * \Delta AvFreq_{i,t} + \\ & \eta_{c4} * AvFreq_{i,t} * \Delta AvFreq_{i,t} + \eta_{c5} * AvOTD_{i,t} + \eta_{c6} * \Delta AvOTD_{i,t} + \\ & \eta_{c7} * AvOTD_{i,t} * \Delta AvOTD_{i,t} + \sum_{k=1}^{11} \eta_{c8k} * MONTH_k + \eta_{c9} * Vol_{it} + \varepsilon_{cit} \end{aligned} \quad (5.12)$$

and,

$$\begin{aligned} \Delta Rev_{it} = & \lambda_{ri} + \eta_{r1} * \Delta Rev_{i,t-1} + \eta_{r2} * AvFreq_{i,t} + \eta_{r3} * \Delta AvFreq_{i,t} + \\ & \eta_{r4} * AvFreq_{i,t} * \Delta AvFreq_{i,t} + \eta_{r5} * AvOTD_{i,t} + \eta_{r6} * \Delta AvOTD_{i,t} + \\ & \eta_{r7} * AvOTD_{i,t} * \Delta AvOTD_{i,t} + \sum_{k=1}^{11} \eta_{r8k} * MONTH_k + \varepsilon_{rit}. \end{aligned} \quad (5.13)$$

Where all variables are as before.

In equations (5.12) and (5.13), the relationship is a non-linear when the parameters of the change variables and the parameters of the interaction differ in sign. Results for the non-linearity analysis are reported in table 5.14.

¹¹ Unfortunately, these analyses can not be compared with the analysis of Banker et al. (2000b) since they do not provide the necessary information.

¹² Ittner and Larcker (1998a) report problems with multicollinearity when they explore this analysis. However, my sample is more than ten times larger than their sample, which mitigates the problems. In addition, multicollinearity works against finding interactions.

Table 5.14: Incremental information content of non-financial performance measures (non-linearities)

Variable ¹	$\Delta Cost_{it}$ (n=864 ²)		ΔRev_{it} (n=837 ³)	
	Coefficient ^{4,5}		Coefficient	
$\Delta Cost_{t-1}$	-0.373	**		
ΔRev_{t-1}			-0.624	**
			(-61.00)	
$\Delta AvFreq_{it}$	-14.812	**	-1.42	**
	(-22.44)		(-28.36)	
$AvFreq_{it}$	-1.310	**	0.314	**
	(-7.99)		(12.00)	
$\Delta AvFreq_{it} * AvFreq_{it}$	8.895	**	0.783	**
	(19.34)		(25.40)	
$\Delta AvOTD_{it}$	-1.028	**	-0.615	**
	(-9.78)		(-10.15)	
$AvOTD_{it}$	-0.02	**	0.003	
	(-3.92)		(1.09)	
$\Delta AvOTD_{it} * AvOTD_{it}$	0.011	**	0.006	**
	(9.66)		(9.47)	
Vol_{it}	0.000	**		
	(34.86)			
Incremental F-test for non-linearities	F=210.95 (p<0.01)		F=323.18 (p<0.01)	
R ² -adj.	0.624		0.370	

¹ The month dummies and intercepts for each area are considered to be nuisance variables, therefore they are not reported.

² Because 4 lags are the maximum in this model, and a change model is used (945-(4*27))=837 observations are available.

³ Because 5 lags are the maximum in this model, and a change model is used (945-(5*27))=810 observations are available.

⁴ Where **, * means significant at the 0.01, and 0.05 levels (two-tailed test).

⁵ t-statistics between parentheses. All results are from the seemingly unrelated regression.

F-tests support that non-linearities exist in the data set for both the cost (F=210.95, p<0.01), and revenue (F=323.18, p<0.01) model. The significant interaction between $\Delta AvFreq_{it}$ and $AvFreq_{it}$ of 8.895 (t=19.34, p<0.01) means that at low (high) worker satisfaction levels a further decrease (increase) leads to a higher increase in future costs than at a higher (lower) worker satisfaction level. This implies that it becomes increasingly more difficult to improve worker satisfaction at high levels. In addition, at a high on-time delivery level ($AvOTD_{it}$) a further increase in on-time delivery leads to more increases in future costs than at a lower on-time delivery level, represented by the significant interaction of 0.011 (t= 9.66, p<0.01). This suggests again that it becomes increasingly more costly to increase on-time delivery.

Again results for the Rev_{it} -model are puzzling at least. At a low worker satisfaction level a further decrease in worker satisfaction leads to a higher increase in future revenues, represented by the significant interaction of 0.783 (t=25.40, p<0.01). In addition, at a high level of OTD an increase in OTD leads to a higher increase in

future revenues than at a lower OTD, represented by the significant interaction of 0.006 ($t=9.47$, $p<0.01$). These results are completely counter-intuitive and unexplainable.

In sum, the tenor of these results suggests that the relationship between non-financial measures and future financial variables is non-linear. For the cost model it becomes increasingly more difficult to improve the non-financial measures when they are already at a high level. For the revenue model it is more difficult to explain the results.

5.5 Discussion and summary

The accounting literature is ambiguous whether non-financial performance measures have relative or incremental information content, or both, to financial measures for predicting future financial performance. Therefore, in this chapter I directly test the claim that is advocated in the popular management literature, that non-financial performance are better predictors for future financial performance than lagged financial measure. In addition, I add to the growing literature that considers incremental information content of non-financial performance measures.

I find that the non-financial measures do not have more relative information content than lagged financial measures. Additionally, I find no support that relative information content of non-financial measures increases compared to lagged financial measures when more lags are used in the models. In contrast, the change models indicate that a change in the financial measure has more information content for predicting change in future financial performance than the non-financial measures.

The non-financial measures however have incremental information content beyond the lagged financial measures for both future cost and future revenues. More specific, I find that the non-financial measures worker satisfaction and on-time delivery have incremental information content beyond the one period lagged costs and revenues for both costs and revenues. Higher worker satisfaction leads to more future costs, but also to more future revenues. The on-time delivery measure has a positive impact on both costs and revenues. The incremental information content of the two lagged non-financial measures beyond the lagged financial measures is roughly one-seventh beyond costs and one-fifth beyond revenues. The lag search procedure indicates that the individual non-financial measures have different lags for cost and revenues. Finally, I find that the incremental information content of the non-financial measures is nonlinear. For the cost model it becomes more costly to improve non-financial measures at higher levels. Results for the revenue model are counter-intuitive. These results suggest that at higher levels of the non-financial measures a further increase in the non-financial measures has a higher benefit than at lower levels of the non-financial measures.

The study suffers from a number of limitations. First, the results are found in one particular organization and therefore any generalizability of the results should be interpreted with caution. Second, for the lag search I assume for reasons of simplicity that the lag is the same for each cross-section. This is not necessarily true of course. Another limitation of the lag search is that it assumes that non-financial measures are indicators for financial measures. However, improving non-financial measures probably will probably cost money first. Therefore, statistical methods that take the bi-directional causal nature of the relationship into account should be considered to explore the relationships between financial and non-financial measures (Luft and Shields, 2001).

Third, I only examine the information content of non-financial measures. To assess whether non-financial measures are useful for contracting purposes it is also important to consider the sensitivity and noise of the measure (Banker and Datar, 1989).

The study suggests a number of interesting avenues for future research. First, the results found should be corroborated in different settings and with different non-financial measures. Second, further research could consider what factors explain differences in relative and incremental information content. Third, I find that higher worker satisfaction measure leads to higher future costs whereas higher on-time delivery measure leads to less future costs. Therefore an interesting research question would be which characteristics of non-financial measures leads to either higher or lower future financial performance. In addition, I find different lags between different non-financial measures and financial measures. Further research could consider what factors might explain these different lags. Fourth, it is interesting to explore what the difference between relative and incremental information contents means for contracting. The theoretical literature suggests that relative information content, i.e., the congruence of different measures, is an input for the weight placed on the different measures¹³ (Datar et al., 2001), whereas the incremental information content of measures gives insights in which measures have potential for using in contracts. However, little empirical evidence exists about these issues.

¹³ Assuming risk-issues are constant.

Chapter 6

Performance effects of including non-financial measures in the contracts of managers

6.1 Introduction

This chapter addresses the second research question, i.e., do managers perform better when non-financial measures are added to their performance evaluation system? As discussed before (see section 1.1) the accounting literature is full of suggestions of management accounting innovations. However, there are hardly any studies that systematically evaluate such innovations. The study in this chapter is an empirical account of such an innovation in one company. The data used in this study is defined in section 4.2. In addition, the background information about the company given in chapter 3 is used to explain the results.

Earlier research that considers the relationship between non-financial performance measures use and performance are often cross-sectional and show inconclusive results (Abernethy and Lillis, 1995; Ittner and Larcker, 1995, 1997; Chenhall, 1997; Perera et al. 1997; Van der Stede, 2001). This study differs from other studies in a number of ways. First, in contrast with most studies, this study has a longitudinal research design. This is often stated as a useful improvement to consider the long-run performance effects of an innovation (Perera et al., 1997). This study differs in a number of aspects from the only other study that uses a longitudinal design. One limitation of the study of Banker et al. (2000b) is that next to the added non-financial measures the company also increased the maximum bonus of the managers.¹ This effect might contaminate the impact of the change in the PMS. Therefore the current study is a purer test of the value of adding non-financial measures into the contract of managers. Second, in this study different non-financial measures are added to the contract. In this study the organization added a worker satisfaction measure to the contract, whereas Banker et al. (2000b) consider the impact of adding a customer satisfaction measure to the contract.

The chapter is structured as follows. Section 6.2 discusses the framework used that guides the research question. Section 6.3 explains the research design used to answer

¹ The existing bonus before the change in the PMS was below the industry average (Banker et al., 2000b, p. 69).

the question. Section 6.4 outlines the research methods. Section 6.5 gives the specification for the empirical analysis, and discusses some statistical considerations. Section 6.6 gives the empirical results. Finally, section 6.7 summarizes the results, and gives limitations, and directions for further research.

6.2 Framework

6.2.1 Introduction

Results from formal agency models, more specific from principal-agent models, can guide the research question addressed in this chapter. A second theoretical justification why the change in the PMS might be beneficial for the organization is the notion of complementarity (Milgrom and Roberts, 1995). This assumes that a particular strategy can only be successful when other systems of the company are congruent with this strategy. For example, Wruck and Jensen (1994) suggest that a successful implementation of a TQM-system require changes in the decision rights, performance measurement system, and reward system. Milgrom and Roberts (1995) argued that complementarities means that the returns of A increase when the level of B is higher. In the current research context, the positive effects of using a quality strategy in the company (see chapter 3) increase when the PMS is congruent with this strategy.

Section 6.2 is structured as follows. First, agency theory is introduced. Second, the role of principal-agent models in explaining management accounting procedures is described. Section 6.2.4 explains the value of adding measure to a contract. Next, limitations of the principal-agent models are enumerated. Finally, expectations for the impact of the change in the PMS are derived.

6.2.2 Agency theory

Agency models explain how features of information, accounting, and compensation systems influence incentive problems, and how these incentive problems influence the design of information, accounting, and compensation systems (Lambert, 2001). It thus assumes that there are conflicts of interest², motivational problems, and explains how mechanisms can mitigate these problems.

The agency model compares information, accounting, and incentive systems when they are used optimally with respect to the amount of incentive problems. This suggests that no performance improvements can be gained by adopting certain accounting systems³. However, this state of equilibrium is often not realistic in organizations as they are in a process of constant updating their systems, reacting on exogenous shocks in the internal and external environment of the organization (Milgrom and Roberts, 1992; Ittner and Larcker, 2001). Baiman (1990, p. 567) suggests that one strategy to increase the contribution of agency models to management accounting is “to concentrate less on deriving optimal compensation contracts, and concentrate more on more easily observed aspects of the firm”. In this

² Causes of conflicts of interest are, among others, effort aversion of the agent, different time-horizons between the principal and the agent, and risk-aversion of the agent (Lambert, 2001).

³ This also implies that under this condition, “..empirical researchers should not even attempt to explain organizational performance because any statistical significant coefficient on the managerial accounting choice will only occur because of measurement error, misspecification of functional form, inadequate set of exogenous controls, etc. (Ittner and Larcker, 2001, p.398)”.

dissertation, with the change in the PMS examined the organization strives for optimality. Consequently, this is tested in the empirical analysis (Banker et al., 1996a; Banker et al., 2000b).

Although the agency models have attractive features for studying management accounting procedures, the model cannot explain why, for example, rewards and punishments are asymmetrically distributed, and pay systems often lead to the same rewards for many employees (Baker et al., 1988).

6.2.3 The principal-agent problem and management accounting procedures

A principal-agent relationship exists when the principal delegates some of his responsibilities to others, the agents⁴. Further, the goals of the principal(s) are not the same as the goals of the agent(s), and the principal cannot, or it is too expensive, completely observe (monitor) the behavior and actions of the agent. Hence, there is information asymmetry between the principal and agent, about the effort chosen by the agent and the possible states of nature. Since the principal is assumed to be risk neutral he could bear all the risk and give the agent a fixed wage. However, in that case a work averse agent has no incentive to work. In contrast, production would be optimal if the principal sold the firm to the agent for a fixed fee. In that case the agent would bear all the risk. However, the risk-averse agent is asking for an extra compensation for this risk bearing.

Thus, the principal-agent model can be characterized as a trade-off between the optimal risk bearing by the agent for incentive purposes, and the risk premium the agent is asking. The principal-agent problem exists when there is information asymmetry between the principal and agent, and the agent is risk and work averse.

Principal-agent theory is one of the few coherent economic theories that addresses managerial accounting procedures and poses managerial accounting questions (Baiman, 1990). The model is studied to explain the role of managerial accounting procedures, such as budgeting, monitoring, cost allocation, and transfer-pricing. Based on the principal-agent model one would expect to find these managerial accounting procedures only in situations in which individuals (or organizations) would benefit from their use (Baiman, 1990). In addition, one could use the principal-agent model to explain the usefulness of two competing systems in an organization. If the second system better mitigates the agency problems than the first, the second system is to be preferred.

In this study, the principal-agency model is used in the context of a changing performance measurement system. In the old system managerial performance is measured, evaluated and rewarded mainly on financial performance measures, while in the new system managerial performance is evaluated and rewarded based on relatively more non-financial performance measures⁵. Observing outcome variables, which are indicators of the agents' actions, assesses preference for the two systems.

⁴ The description of the principal-agent problem is based on Baiman (1990).

⁵ Principal-agent models explore the incentive function of performance measures use. The incentive function is a composition of both the evaluation and reward function (Baiman, 1990).

6.2.4 Value of non-financial performance measures in the multi-task principal-agent framework

This section describes the value of using additional measures in the contract in the multi-task principal-agent framework. The context described is the value of non-financial measures that are added to the financial measures in the contract.

The primary distinction of a multi-task principal-agent model is the premise that the agent performs more than one task or that the task is multi-dimensional (Holmstrom and Milgrom, 1991)⁶. This raises the issue of incomplete measures in agency models, i.e., incongruence. Next to the functions of allocating risk between the principal and agent, and inducing effort, in a multi-task principal-agent model the contract directs the choice of the agent to allocate his attention among the different tasks, or dimensions of the task. This means that an increase in the agents' compensation tied to one task will reallocate his attention to this task, and lessen the attention to other tasks.

The two most important characteristics of a performance measure are its congruence with the principal's expected gross payoff, and its noise, that is uncontrollability, of the measure. Non-congruent measures lead to sub-optimal effort allocation, whereas noise of the measure leads to sub-optimal effort intensity (Feltham and Xie, 1994). The value of additional performance measures over an existing measure differs in both cases. First, if the existing measure is non-congruent, additional measures increase the action set of the agent and therefore makes the behavior of the agent more congruent with the principal's expected gross pay-off. Second, additional measures can reduce risk imposed on the agent due to noise, i.e., uncontrollability, in the existing performance measure (Feltham and Xie, 1994).

The effects of evaluating managers (agents) only on a measure which does not appropriately incorporate the economic consequences of all relevant activities are often described in the multi-task principal-agent literature (Hemmer, 1996). This can lead to a short-run orientation or have negative side-effects on other departments (externalities). An often used example is hard selling. Sales people can improve their sales measure by two strategies. First, the salesman aggressively sells products to his customers. This strategy can in the long-run lead to unsatisfied customers. The sales measure probably rises in the short-run, but may deteriorate in the long-run when customers feel unsatisfied with the product. A second strategy is to provide extra service to the customers, which will make them more satisfied, and leads to more sales in the future. The principal can induce the second strategy, which is more congruent with his goals, by contracting the agent on a sales and a customer satisfaction measure.

The case of risk reduction can formally be expressed in the following way. If measure x is a function of effort (a) and noise (ε_x), that is $x=a+\varepsilon_x$, and y is only a function of noise, that is $y= \varepsilon_y$, then the value of adding measure y to the contract is only 0 in the case where the correlation (ρ) between x and y is 0 (Feltham and Xie, 1994). In practice the case of risk reduction is seen with relative performance evaluation. The performance of agent two says something, i.e., $\rho(x,y)\neq 0$, about the performance of agent one, even he it cannot influence the measure itself. In the case of the salesman, if customer satisfaction is an indicator of sales, that is, financial and

⁶ Another distinction is that in a multi-task principal-agent model the design of the job becomes a relevant issue. When a function consists of more than one dimension, these dimensions can be aggregated into one person or in different persons. An example is separating the production and quality control of production workers into two persons, or incorporating it into one.

non-financial measures are correlated, adding non-financial measures to financial measures has value.

In sum, adding non-financial measures to financial measures in the contracts of managers can lead to a better allocation of effort, and to reduce the overall risk of the manager.

6.2.5 Limitations of the principal-agent model

In this section, I discuss a number of limitations of the principal-agent model, and the way my empirical study addresses these limitations. The principal-agent model has been the subject of severe criticism regarding the assumptions underlying the model, the outcomes of the model, and the simplicity of the model (Baiman, 1990). This makes it difficult to perform a direct test of the model.

A criticism regarding the outcomes of the model relates to Holmstrom's result, called the informativeness principle, that information about the effort of an agent needs to be in the contract if it gives incremental value over the other signals in the contract (Holmstrom, 1979). This leads to the outcome that many (infinite) measures should be used in the contract. Together with the assumption that contracts are costless⁷, the result of the principal-agent model gives complicated and complete contracts. In practice, simple (linear) and incomplete contracts are often found (Baiman, 1990). Reasons for this observation are higher costs, both in time and money, of maintaining the information system, and cognitive limitations of managers⁸, which leads to dilution of attention to each individual measure when contracts are extended with additional measures⁹. For example, in a Perrin Towers survey under 60 companies that adopted the Balanced Scorecard, the large number of measures was one of the problems managers indicated (Ittner and Larcker, 1998b). Therefore, Holmstrom's result is probably only true for the region where the incremental value of an extra signal about the effort of the agent is less than the extra costs of gathering data for the signal. Therefore, added value of additional performance measures that are indicators of managerial effort, in practice is expected to be an inverted U-curve.

In the current study, the evaluation and rewarding of managers is based on a limited number of performance measures that are important for the organization (see table 3.1). This suggests that the added value of adding measures in the contract of the managers of this organization may still be in the increasing part of the inverted U-curve.

6.2.6 Effect of change in PMS on financial and non-financial measures

Defining performance is difficult in every study. Perera et al. (1997) remark that:

“Performance is a complex variable with a multiplicity of factors contributing to the level of global performance at any point in time” (Perera et al., 1997, p. 569).

⁷ In a principal-agent model extra measures cannot have a negative value because the principal can always assign zero incentive to the measures (Feltham and Xie, 1994).

⁸ Miller (1956) found that individuals can store only 7 plus or minus 2 information elements in their working memory. Empirical evidence shows that for non-management jobs three till five measures is the optimum. Afterwards, benefits decline (McAdams and Hawk, 1994).

⁹ Baiman (1990) suggests incorporating a contract design cost within the agency model, but admits that the form of such a measure is not well understood.

Therefore, for assessing the effectiveness of the use of relatively more non-financial performance measures to evaluate managers, both financial and a number of for the organization important non-financial performance measures are used. Becker and Gerhart (1996, p.791) argued that effectiveness measures should be “natural and meaningful measures” that have inherent meaning in the context of the research problem. The effectiveness measures used in this study are the measures from the contracts of the area managers, and are therefore indicators of managerial performance. All dimensions of performance partly correlate with the manager’s action, i.e., effort level. Although the study in this chapter is not a formal test of the principal-agent model, insights from the model can guide the research question.

In a multi-task principal-agent model the weight placed on a measure of a task directs the amount of effort of the agent to the task (Feldham and Xie, 1994). Hence, adding additional measures to the contract, in this case non-financial measures, increases the effort allocated to the tasks of the added measures. Additionally, based on self-selection the new contract will attract managers that are better in executing the tasks that are emphasized relatively more (Banker et al., 2000a). Therefore, the added non-financial measure worker satisfaction is expected to increase when added to the contracts. There is little empirical evidence of the impact of adding measures in compensation plans in longitudinal studies. Banker et al. (2000b) find a positive impact of the added non-financial measure customer satisfaction after its inclusion in the bonus plan. Symons and Jacobs (1995) find that operational performance improved after the inclusion of a TQM-based rewards system for production workers was installed. However, Pearce et al. (1985) find no impact of adding 4 organizational performance measures in a compensation plan.¹⁰

In contrast, principal-agent theory states that the tasks for the measures that were in the contract already get relatively less emphasis, and are therefore expected to decrease. This would lead to the expectation that the on-time delivery measure that was already in the contract is expected to decrease. However, when increased worker satisfaction has a positive impact on the on-time delivery measure this might lead to an increase in on-time delivery. Therefore, no direction is formulated for the impact of the change in the PMS on the on-time delivery measure.

Again, principal-agent theory states that the tasks for measures that were in the contract already get relatively less emphasis, and are therefore expected to decrease. However when the non-financial measures are leading indicators for the financial measures the financial measures might also increase after the change in the PMS. In addition, the multi-task principal-agent model assumes that the weight of the existing measure becomes less when a new measure is added to the contract.¹¹ However, some empirical evidence shows that when relatively more non-financial measures are used, this does not mean that less emphasis is placed on financial measures (Govindarajan and Gupta, 1985; Carr et al., 1997; Van der Stede et al., 2001).

The impact of the change in the PMS might have different effects for either cost or revenues. Next to the effect that less emphasis is given to measures that were in the contract already, improving non-financial measures might lead to either more or less future cost (see section 5.2.2). Therefore, the change in the PMS has a differential

¹⁰ In this study like in Banker et al. (2000b) study, before the change in the PMS there was no bonus awarded to the managers and therefore there are two simultaneous changes in the PMS.

¹¹ Theoretically, this is true under most conditions. However when the (normalized) co-incongruence and correlation between the two measures are highly negative, it is possible that the weight on the existing measure increases after a measure is added, because the added measure filters noise and offsets distortions of the existing measures. (Banker and Theveranjan, 2000).

effect on the cost measure and it is not clear which effect, i.e., the “less weight” versus “leading indicator” is dominant. The empirical evidence from Banker et al. (2000b) shows that the inclusion of non-financial measures in the contracts of managers did not lead to a decrease or increase in the cost measure in their context.

For the revenue measures, improved lagged non-financial measures are expected to have a positive impact on revenues (see the empirical results in chapter 4). In contrast, principal-agent theory would predict that existing measures in the contract get less weight when the contract is extended with extra measures. Since the number of items in the contract of the managers increased only from three to five, and therefore the attention for each measures is not diluted too much, I suggest that the “leading indicator” argument is dominant over the “less weight” argument. Thus, I expect that revenues do increase after the inclusion of the non-financial measures in the contract of the managers. The empirical evidence in Banker et al. (2000b) shows that the revenues do indeed increase after the inclusion of the non-financial measure customer satisfaction in the contract.

Table 6.2 summarizes the expectations derived from the framework used in this study.

Table 6.2: Summary of expectations of adding non-financial measures to the contracts of managers

Impact of change in PMS	Expectation¹
Non-financial measures ²	
Change in PMS → Worker satisfaction	+
Change in PMS → On-time delivery	?
Financial measures	
Change in PMS → Costs	?
Change in PMS → Revenues	+

Notes:

¹ Where a + means that the change in the PMS is expected to have a positive impact on the measure, a ? means that the change is expected to have a differential effect on the measure.

² Only the worker satisfaction measure was added to the contract.

6.3 Research design

6.3.1 Introduction

This section explains the research design for this study. First, the research method of experimentation, more specific the quasi-experimental design, is described. Next, a short description of the interrupted time-series design is given. Afterwards, potential problems, i.e., threats to internal validity, with this method are enumerated.

6.3.2 Quasi-experimental design

Experiments can be characterized by their structure and function (Cook and Shadish, 1994). The structure of an experiments consist of a) a sudden intervention, b) knowledge of when the intervention occurred, c) one of more post-intervention outcome measures, and d) some form of counterfactual, that is a base-line, to compare the after intervention outcome measure with.

The function of an experiment is to test causal hypotheses. Causal relationships are dependence relationships between two or more variables in which the researcher clearly specifies that one or more variables “cause” or create an outcome represented by at least one other variable (Hair et al., 1998). Cook and Campbell (1979) give three

conditions that have to be met before confidence in a causal relationship can be derived. First, there must be a temporal ordering, where the cause comes before the effect in time. Second, the cause and effect have to co-vary.¹² Third, there are no plausible alternative explanations for the observed effect.¹³

Social experiments, or the more familiar name field experiments, differ from laboratory experiments in that the subjects are less isolated from their environment, and intervention procedures are less standardized (Cook and Shadish, 1994). Although this is considered an advantage of the method, because results found are considered to be more generalizable (Cook, 1979), it is also a weakness. Especially the third condition of causality is hard to meet, because an unlimited number of factors, next to the treatment can be the reason for the effect observed. This leads to a number of potential alternative explanations for the effect found.

Social experiments can be classified as either randomized experiments or quasi-experiments. Quasi-experiments differ from randomized experiments in the selection criteria of the assignment of the treatment over the subjects. In randomized experiments the subjects that get the treatment are randomly chosen. In contrast, in quasi-experiments the treatment is assigned to subjects based on self-selection or administrative procedures (Cook and Shadish, 1994). This non-randomness is a serious threat for the causal inference of a relationship. For example, the treatment can be based on self-selection, that is more qualified subjects get the treatment, whereas, less qualified end up in the control group. However, researchers hardly ever get an opportunity to execute, or observe, a randomized social experiment. Therefore most experiments are non-random.

6.3.3 Interrupted time-series design

From the large family of quasi-experimental designs the interrupted time-series design is considered to be one of the strongest (Marcantonio and Cook, 1994). The relative strength of the interrupted time-series design compared with, for example, a simple before-after design is that in the interrupted time-series design a trend that already exists in the time-series can be controlled for.

Figure 6.1: Before-after design

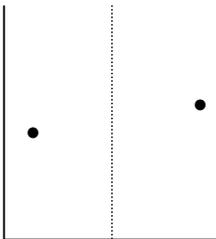
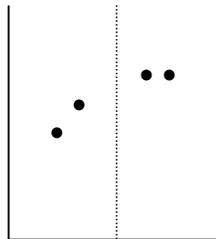


Figure 6.2: Interrupted time-series design



For example, in figure 6.1 a simple before-after test would indicate that the intervention had a positive impact. However, figure 6.2 indicates that the time-series increased already before the intervention, and therefore the intervention had no positive effect.

The interrupted time-series design is especially strong when a sudden intervention took place, and a rapid effect is expected or the delay before the intervention has an

¹² Called statistical conclusion validity (Cook and Campbell, 1979).

¹³ Called internal validity (Cook and Campbell, 1979).

effect is known (Marcantonio and Cook, 1994). The basic interrupted time-series design can be represented schematically in the following way:

$$O_{-n} \dots O_{-4} O_{-3} O_{-2} O_{-1} X O_1 O_2 O_3 O_4 \dots O_n$$

Where:

O_{-n} through O_{-1} are observations before the treatment,
 O_1 through O_n are observations after the treatment, and
 X is the treatment.

In this design the post-treatment observations are compared with the pre-treatment observations, i.e., the counterfactual, to estimate the impact of the intervention. This basic design can be extended with additional features, for example with a control group, additional outcome variables, more than one or switching interventions, and more explanatory variables in the analysis that explain the outcome measures. In the study in this chapter, the post change outcome measures are compared with the trend in the time-series of the outcome measures before the change.

6.3.4 Threats to internal validity

The interrupted time-series design is subject to a number of threats to internal validity (Cook and Campbell, 1979). Internal validity deals with ruling out alternative explanations for the causal inference, and is the most discussed type of validity in a quasi-experimental design. Cook et al., (1990) give a list of threats to internal validity specific for the interrupted time-series designs. These threats are 1) maturation, 2) the cyclical nature of time-series, 3) main effect of history, 4) instrumentation, and 5) selection effects. Cambell argued that only plausible threats need to be ruled out, but admits that plausibility is a slippery concept (in, Cook and Shadish, 1994).

First, in the case of maturation another trend, independent of the treatment, started already before the intervention. This threat is for example due to learning experience, or just ageing. I control for maturation by assuming that the trend in the time-series before the change in the PMS persists after the change.

Second, a seasonal effect may hide the effects of the treatment. Identifying the source of seasonality, for example, monthly or quarterly, can control for this effect. In this study, I use dummies for each month to control for the cyclical nature of the time-series.

Third, a main effect of history exists when alternative causes started at the same time of the treatment. Possible solutions for this effect are the use of a control group, and a record of all plausible effect-causation events. The problem of history is mitigated when the interval between two observations is smaller, since the probability that two or more causes started at the same moment is decreasing. In this study, main effects of history are qualitatively examined, by interviews with different managers from the organization, or from theory.

A fourth threat is changes in recordkeeping practices, called instrumentation during the measurement period. This problem reveals itself in a changed definition of the time-series over time, or a more than normal emphasis on following the measure. Since an intervention is often a process of re-organizing existing procedures, the instrumentation problem is often a problem. Instrumentation is addressed in this study in a number of ways such as interviews with managers, investigating several internal documents, and examining time-series visually for sudden breaks.

The last threat is selection and mortality. A selection effect is an effect that exists because the analyzed units have different attributes and therefore may react differently on the intervention. A related effect is mortality. In that case, subjects with bad attributes for the specific treatment drop out. Possible solutions for these threats are the measurements of units that have complete time-series, or a background analysis of all units to identify attributes of the subjects. In this study, no control is used for the selection effect. In the principal-agent paradigm both the effort effect and selection effect are determinants for performance effects (Banker et al., 2000a). Hence, a control for the selection effect would absorb a part of the effect we are interested in.

The threats to internal validity specific for the interrupted time-series design and the way this study addresses these threats are summarized in table 6.3.

Table 6.3: Threats to internal validity

Threats to internal validity ¹	This study
Maturation	Is assumed to be controlled for through the use of a linear increasing base-line in the specification
Cyclical nature of time-series	Use indicator variables for each month
Main effect of history	Qualitative examination by interviews
Instrumentation	Interviews, visual examination of time-series, and internal documents
Selection effect	No controls since selection effect is part of the theory

Notes:

¹ These threats are specific for the interrupted time-series design (Cook et al., 1990).

6.4 Research methods

6.4.1 Introduction

This section reviews different methods to analyze time-series and assess evaluation questions. First, two different research methods to analyze the panel data set, i.e., time-series analysis and time-series regression, are compared. Afterwards alternative specifications to estimate the impact of the change in the PMS are examined. Finally, evaluation studies in accounting research, and their research methods are reviewed.

6.4.2 Time-series regression versus time-series analysis

Autoregressive integrated moving average (ARIMA) modeling and regression approaches are two different approaches to analyze time-series¹⁴. The most important difference is that regression models are built from prior theory and/or research, whereas ARIMA models are build from the time-series itself (Orwin, 1997).

A problem with the regression method is that there is often no theory to explain the time-series completely. A statistical problem is that time-series often contain autocorrelated observations. When autocorrelation is present an observation is

¹⁴ In addition, there are more approaches available to analyze time-series. It is also possible to use a combination of the regression and ARIMA approach.

explained by one or more of its former observations. Although neglecting autocorrelation in regression models does not bias coefficients, they are not efficient (Greene, 2000). In the more general case of positive autocorrelation the standard deviation is deflated, which leads to more significant coefficients (Mohr, 1995).¹⁵

The ARIMA, or Box-Jenkins, model mitigates the problem of autoregression. It differs from the better known regression model, in that a time-series is explained in terms of autoregressive and moving average¹⁶ processes that characterize the series itself. In addition to “whitening” the time-series¹⁷, i.e., removing any recurring systematic components from the data, a transferfunction is used to assess the impact of an intervention into the series (Orwin, 1997). These transferfunctions can take different forms depending on the expected impact of the change.

A disadvantages of ARIMA-modeling is that the removal of the trend by differencing, and the estimation of the autoregressive and moving average processes are a-theoretical, i.e., it is driven by the data and can remove the effects of unmeasured variables which can interact with the intervention (Pearce et al., 1985).¹⁸ Further, because ARIMA modeling is an empirical method it needs relatively long time-series. McCleary (1980) gives as a rule of thumb a minimum of 50 observations.

The predominant factor why time-series regression, instead of time-series analysis, is used in this dissertation is that only 36 observation for each time-series is available. Small time-series lead to relatively high standard errors of the parameters. Therefore, the time-series regression approach is preferred.

6.4.3 Impact assessment

Cook and Campbell (1979) discuss three elements that drives the proper specification to estimate a possible impact, i.e., 1) the shape of the impact, 2) the type of the impact, and 3) the permanence of the impact.

First, the appropriate specification to estimate the impact of an intervention depends on the expected impact of the intervention. An intervention can lead to a 1) change in the mean of the time-series (a structural shift), 2) a change in the slope of the time-series (a gradual shift), and 3) a change in the variability round the mean or a change in the seasonal pattern, or 4) a combination of 1, 2 and 3 (Cook and Campbell, 1979).

Second, the impact can be instantaneously or delayed. Estimating the impact of an intervention is easier when the exact date of the intervention is known and the expected effect comes rapidly or has a known delay (Marcantonio and Cook, 1994). Unfortunately, in management accounting research there is little theory that guides the expectation of such a possible lag.

¹⁵ Cook and Campbell (1979) give an indication of the impact of autocorrelation on significant testing.

$$t_{\text{test}} = t_{\text{real}} * (1 + \theta) / (1 - \theta).$$

Thus, for example in the case of an autocorrelation coefficient (θ) of 0.5. The T_{test} is inflated by 225%. In that case t-statistics of 4.5 are not significant.

¹⁶ Moving average processes are positive or negative random shock that influence the time-series.

¹⁷ For a complete overview of the method is referred to McCleary and Hay (1980).

¹⁸ Although the a-theoretical solution of the problems is a standard critic, for example in Harvey and Durbin (1986), controlling for a trend in the data that existed already before the event, transforming observations for autoregression, and controlling for a seasonal pattern in time-series regression has the same characteristic.

The third element is the permanency of the effect. Cook and Campell (1979) argued that most effects decay over time even if the program or new system is still in place. The opposite possibility of an impact that grows over time is less often observed.

The different possible shapes of an expected impact all use different specifications. The simplest specification to assess the impact is a change in the mean of the time-series:

$$y_t = \alpha + \beta_1 * D_t + \varepsilon_t . \quad (6.11)$$

Where:

D_t is 0 before the intervention and 1 after the intervention¹⁹, and y_t is the outcome variable.

In equation (6.11) the outcome variable after the implementation is compared with the outcome variable before the implementation.

Equation (6.12) is a more sophisticated specification that controls for a trend in the time-series that already existed before the intervention:

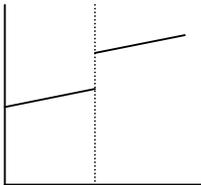
$$y_t = \alpha + \beta_1 * T_t + \beta_2 * D_t + \varepsilon_t . \quad (6.12)$$

Where:

T_t is a trend in the data measured by 0 in period 0, 1 in period 1, and 2 in period 2, etc., and other variables are as before.²⁰

The rationale behind equation (6.12) is that when a trend existed already before the intervention, this cannot be explained by the intervention itself.²¹ This is a control for the maturation treat to internal validity. β_2 estimates the impact of the intervention. This specification assumes a direct abrupt effect of the intervention. Equation (6.12) is graphically represented in figure 6.3.

Figure 6.3: Graphical representation of an abrupt shift in the time-series



An alternative specification that allows the effect to be abrupt or gradual is:

$$y_t = \alpha + \beta_1 * T_t + \beta_2 * D_t + \beta_3 * D_t * T_t + \varepsilon_t . \quad (6.13)$$

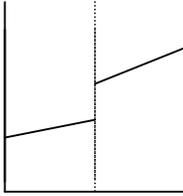
¹⁹ This specification is a t-test on the average mean before and after the change.

²⁰ This assumes a linear trend, however other specifications are also possible.

²¹ This specification is a problem in the case of dependent variables measures estimated in percentages, for example in percentage of rejections. In the case of reaching 100% it is more difficult to improve the measure.

In equation (6.13), β_3 gives a change in the slope of the time-series, whereas β_2 still gives a shift in the intercept of the time-series. Equation (6.13) is graphically represented in figure 6.4.

Figure 6.4: Graphical representation of a gradual and abrupt change in the time-series



A change in the variability of the time-series might be expected when organizations are better able to control the process, and therefore the variance of the time-series decreases.²²

The methods described above to estimate the effect of a certain treatment can all be extended to capture other factors beside the intervention that can influence the time-series.

6.4.4 Impact assessment in accounting research

Few studies in management accounting literature estimate the impact of a change in the PMS through an interrupted time-series design. In Banker et al. (1996b) the impact of an incentive plan based on sales, for sales consultants in a retail organization is explored. This is the only study where the treatment is randomly assigned to the cross-sections. Therefore the counterfactual of the cross-sections where the treatment is implemented are the cross-sections where it is not implemented. They estimate the impact of the change in the incentive system through considering models with either a shift in the trend of the time-series, a shift in the intercept, and both, that is specification (6.13) is used.

Banker et al. (1996a) use the following specification to estimate the impact of the change under study (specification given without control variables):

$$Out_{it} = \alpha_{it} + \beta_1 * preT_t + \beta_2 * postT_t + \varepsilon_{it} . \quad (6.14)$$

Where:

Out_{it} = the outcome measure of cross-section i at period t ,

$PreT_t$ = a variable that is 0 after the change and has the number equal to the period before the change,

$PostT_t$ = the variable is 0 before the change and has the number equal to the period after the change.

The $PreT_t$ variable controls for a trend in the time-series that is independent of the treatment. Coefficient β_2 of the $PostT_t$ variable is the indicator for the effect of the change. Because β_1 was not, and β_2 was significantly different from zero this led

²² Wruck and Jensen (1994) also argued that TQM organizations use a different way of setting targets. Next to setting a target for the level of a variable, these companies often stress a limited variance of a measure to increase its predictability.

them to the conclusion that the change had a positive impact. An F-test that β_2 is statistically different from β_1 would be the proper test if the change in the PMS had an impact. This specification does only ascertain whether there is a change in the slope of the time-series and not whether there is a change in the intercept of the time-series.

Banker et al. (2000b) use ARIMA modeling to analyze the impact of the change in the PMS. Banker et al. (2000b, p. 77) state that:

“a significant change in the *slope* or *level* predictable from the model describing the time-series indicates that the intervention event had an effect on the dependent variable” (italics added).

However in the econometric models they only use a dummy, being 1 after the change and 0 before the change, and therefore they only consider a change in the level of the time-series. The reason that the specification in Banker et al. (2000b) is not used in this study is because of the relative short time-series, i.e., 36 periods.

Finally, Emsley (2000) tests the impact of the change under study by a t-test between the average performance after and before the change. Statistical controls used are a test that there was no trend already before the change, and a test of autocorrelation. Further, only a qualitative indication of disturbing factors is given.

In sum, the accounting research literature uses a wide variety of methods to estimate the impact of an intervention. This can be explained by the lack of theory available that should guide such a specification.

6.5 Specification

Cook and Campell (1979) argued that the impact of a program can lead to a change in the intercept (a structural shift), slope (a gradual shift), variance of the time-series, or a combination of these. Ideally, theory should guide the choice of the appropriate specification. Unfortunately, management accounting theory gives no guidance for a possible performance effect of the change in the PMS. Since there is no theory what kind of effect to expect it could be argued that this should be an empirical question, i.e., use a specification that allows either a change in the slope, and a change in the intercept. However, such a specification, comparable with equation (6.13), is subject to multicollinearity problems between the parameter that estimates the change in the slope of the time-series ($D \cdot T$), and the baseline (T) against which the change is compared. The Variance Inflation Factors (VIF) of these parameters are above 40, where a rule of thumb indicates that VIF's above 20 are problematic (Hair et al. (1998). Therefore, I used the structure of equation (6.12) to estimate the impact of the change in PMS. In this specification, the coefficient of the variable D_t estimates the change in the intercept of the time-series, that is a structural impact. This specification assumes that the impact of the change in the PMS starts immediately and is permanent over the time period of the study.

The variable T_t is used to control for a trend in the time-series that existed already before the change in the PMS. This so-called maturation effect controls, for example, for the larger emphasis on quality after the business unit Letters separated the two functions in 1995. T_t assumes that possible maturation effects are linear and persist over time, therefore it might overcompensate. This makes the test of the impact of the change in the PMS very conservative. Especially, for ratios, like the quality and frequency measure, it is more difficult to improve them when they are already on a higher level.

Again, I estimate the fixed effects model, that is in all models dummies for each area are used to control for difference in size, leadership style, etc (see also section 4.3.3).

Thus, for the impact of the change in the PMS on $Freq_{it}$ the following equation is estimated:

$$Freq_{it} = \alpha_{fi} + \beta_{f1} * T_t + \beta_{f2} * D_t + \sum_{k=1}^{11} \beta_{f3k} * MONTH_k + \beta_{f4} * Rsat_{it} + \beta_{f5} * IAZ_{it} + \varepsilon_{fi}. \quad (6.15)$$

Where:

- $Freq_{it}$ = the absence-frequency index, measured by the number of sickness reports per 100 working days,
 T_t = a trend, being -1, -2, -3, etc. before the change, and 1, 2, 3, etc. after the change,
 D_t = the moment of the change in the PMS. Where D_t is an indicator that is 0 before the change and 1 after the change,
 $MONTH_k$ = an indicator being 1 for month k and 0 otherwise,
 $Rsat_{it}$ = the ratio of the number of workers on Saturday from the total number of workers,
 IAZ_{it} = an indicator variable that is 1 when more than 20 out of 10.000 people went to a family doctor in one of four geographical regions, and 0 otherwise.

and other variables and indices are as before.

In the model with $Freq_{it}$ as dependent variable there is a control for a seasonal pattern, because the larger part of the work of most employees is outdoor, hence weather conditions influence the measure. This is a control for the threat of internal validity of the cyclical nature of the time-series. Further, in the meta-study referred to before, McShane (1984) found that age and experience moderated the strength of the relationship between worker satisfaction and absence-frequency. Therefore a control is used for workers on Saturday ($Rsat_{it}$), because these are employees with other characteristics. Most often they are students and are therefore younger and have less experience. Additionally, interviews with human resource managers confirmed the idea that the Saturday workers are less often sick.

The impact on the on-time delivery measure is estimated by:

$$OTD_{it} = \alpha_{oi} + \beta_{o1} * T_t + \beta_{o2} * D_t + \sum_{k=1}^{11} \beta_{o3k} * MONTH_k + \varepsilon_{oi}. \quad (6.16)$$

Where:

- OTD_{it} = the ratio of the number of products supplied at the right place within a pre-specified time period divided by the total number of products,
and other variables and indices are as before.

Again dummies for months to control for a seasonal pattern are used. The impact of the change in the PMS on $Cost_{it}$ is estimated by:

$$Cost_{it} = \alpha_{ci} + \beta_{c1} * T_t + \beta_{c2} * D_t + \sum_{k=1}^{11} \beta_{c3k} * MONTH_k + \beta_{c4} * Vol_{it} + \varepsilon_{cit} . \quad (6.17)$$

Where:

Vol_{it} = number of products,
and other variables and indices are as before.

In the $Cost_{it}$ -model, the dependent variable is deflated by CPI_t to control for inflation. The number of products handled of the organization differs considerable in each time period, e.g., Christmas and New Year is an important period for the organization, therefore, dummies are used for each month. The most important driver of costs is the volume of the production, measured by variable Vol_{it} . Whereas this production volume is for a large part exogenous to the area manager, and is used in the target setting progress, it is also used as a control variable.

The impact of the change in the PMS on revenues is estimated by:

$$Rev_{it} = \alpha_{ri} + \beta_{r1} * T_t + \beta_{r2} * D_t + \sum_{k=1}^{11} \beta_{r3k} * MONTH_k + \varepsilon_{rit} . \quad (6.18)$$

Where all variables are as before.

In the Rev_{it} -model again the dependent variable is deflated by CPI_t to control for inflation. Using dummies for each month controls for a seasonal pattern.

6.5.1 Statistical consideration

The Jarque-Bera statistic reveals no large problems with non-normality. The maximum number of cross-sections for one model that had multivariate non-normal residuals was 10 out of 27. Because of these results and together with the large sample size I choose not to transform the data.

Due to the time-series nature of the data autocorrelation can be expected. Although autocorrelation does not bias coefficients it deflates the standard deviations. Surprisingly, based on the Durbin Watson statistics autocorrelation was not a problem. A plausible explanation for this is that the source of autocorrelation is seasonal, and the use of dummies for each month absorbs this.

One advantage of pooled time-series data is the ability to control for individual heterogeneity of firms, countries etc. (Baltagi, 1995). Therefore, the fixed effects model with dummies for each area is used to control for omitted variables that are constant over time.

Again, I use the seemingly unrelated regression to control for correlated residuals between the areas and cross-sectional heteroskedasticity (see section 5.3.3).

I re-estimated all models without observations that had residuals that are more than 3 standard deviation from the mean (Hair et al. 1998). The maximum number of outliers for a model was 8. These outliers do not influence the results reported.

The R^2 -adjusted are high in most models. These are mainly driven by the fixed effects for each area that control for omitted variables, such as size of the area, and indicator variables for the months to control for seasonality.

6.6 Empirical results

6.6.1 Introduction

Since I use the same variables for this analysis as in chapter 5, I refer for the descriptives and correlation matrix for all variables to table 4.2 and 4.3. The next section reports the results of the change in the PMS on the yearly data. Section 6.6.3 gives the impact of the change in the PMS on the non-financial performance measures. Section 6.6.4 gives the results on the financial measures. Finally, the managerial significance of the results is reported in section 6.6.5.

6.6.2 Impact of change in PMS with yearly data

Before I estimate the impact on the change in the PMS in the interrupted time-series design with monthly data I perform a simple before-after design. I execute a multivariate analysis of variance (MANOVA) with the 5 performance dimensions from the contracts of the managers as dependent variables and a dummy that is 1 after the change and 0 before the change. In this analysis I use the number of products handled in a year as a covariate, since this is exogenous for the area manager. The rationale to execute this analysis is threefold. First, it shows the differences between a simple before-after design, and an interrupted time-series design. Second, I have no proxy for the customer satisfaction measure, but can use this performance dimension in the yearly analysis. Third, in the detailed analysis the impact of the change in the PMS is considered individual in each performance measure from the contract. However, the question whether managers perform better after the change in the PMS is an intrinsically multivariate question (Hair, 1998). MANOVA is the proper research method to analyze these multivariate questions.

The descriptives for the yearly data are reported in table 6.3. The results for this MANOVA analysis are reported in table 6.4.

Table 6.3: Descriptives for yearly data

Variables ¹ (n=81 ²)	Mean	Standard deviation	Median	Minimum	Maximum
Cost	664.84	161.65	657.96	338.33	1125.81
Revenues	364.65	173.76	320.19	171.70	876.89
On-time delivery	90.79	3.96	92.10	81.40	97.00
Worker satisfaction	333.96	16.97	336.00	294.00	369.00
Customer satisfaction	334.27	10.12	336.00	300.00	356.00

Notes:

¹ The cost and revenues measures are deflated by CPI to control for inflation.

² 3 years of yearly observations of 27 areas are available (N=3*27=81).

Table 6.4: Impact of change in PMS on yearly data

	Coefficient ^{1,2}	
<i>Multivariate significant tests:</i>		
-Pillais	0.826	**
-Hotellings	4.757	**
-Wilks	0.174	**
<i>Dependent variables</i>		
Cost	-48.316	**
	(-5.19)	
Revenues	-15.055	
	(-0.77)	
On-time delivery	7.243	**
	(15.70)	
Worker satisfaction	-10.224	**
	(-3.02)	
Customer satisfaction	9.500	**
	(4.67)	
<i>Multivariate homogeneity test (Box-M)</i>	p-value	
	=0.722	

Notes:

This table reports an MANOVA-analysis of the impact of the change in the PMS on the 5 performance dimensions from the managers' contracts. The number of product handled in each cross-section, is used as a co-variate. The data of 1995 (n=27) is compared with the 1996 and 1997 data (n=54).

¹ t-values between brackets.

² Where **, * means significant at the 0.01, and 0.05 levels (two-tailed test).

The multivariate significant tests indicate that the change in the PMS had an impact on the complete set of performance dimensions ($p < 0.01$). The cost measure decreased after the change in the PMS with -48.316 ($t = -5.19$, $p < 0.01$). The change had no impact on the revenue measure. The change in the PMS had a positive impact on on-time delivery, and the added measure customer satisfaction. The on-time delivery measure increases with 7.243 after the change ($t = 15.70$, $p < 0.01$). The customer satisfaction increased with 9.500 after the change ($t = 4.67$, $p < 0.01$). Finally, the added measure worker satisfaction measure decreased with -10.224 ($t = -3.02$, $p < 0.01$) after the change.

A critical assumption of the MANOVA analysis is the homogeneity assumption (Hair et al., 1998), i.e., whether the variance is the same before and after the change in the PMS. The Box-M test does not indicate problems with this assumption ($p = 0.722$).

In sum, the analyses show that the change in the PMS had a positive impact on cost, customer satisfaction, and on-time delivery. The impact on the added measure worker satisfaction, however, was negative. A limitation of this analysis is that there is no control for the trend in the time-series, (see section 6.4.3) and consequently the results might differ compared to the detailed analysis from the next sections. Therefore, for the overall conclusions I focus on the results from the interrupted time-series analysis in the next sections.

6.6.3 Impact of change in PMS on non-financial performance

Table 6.5 gives the results of the impact of the change in the PMS on the non-financial measures.

Table 6.5: Impact of change in the PMS on non-financial measures

Variables ¹	Expected sign	$Freq_{it}$ (n=972 ²)	Expected sign	OTD_{it} (n=972 ²)
Impact:		Coefficient ^{3,4}		Coefficient
D_t	+	0.074 (1.57)	?	0.044 ** (7.04)
Controls:				
T_t	?	-0.006 ** (-2.80)	?	0.001 ** (4.84)
$Rsat_{it}$	+	-7.675 ** (-10.06)		
IAZ_{it}	+	0.388 ** (19.22)		
R^2 -adj.		0.813		0.373
DW-statistic ⁵		1.63		1.86

Notes:

¹ The month dummies and intercepts for each area are considered to be nuisance variables, therefore they are not reported.

² 3 years of monthly observations of 27 areas are available.

³ Where **, * means significant at the 0.01, and 0.05 levels (two-tailed test).

⁴ t-statistics between parentheses. All results are from the seemingly unrelated regression.

⁵ Where the DW-statistic means the Durbin-Watson statistic.

β_{f2} of equation (6.15) measures the impact of the change in the PMS on worker satisfaction²³. The significant coefficient of T_t , β_{f1} , of -0.006 ($t=-2.80$, $p<0.01$), indicates that there was a negative trend in $Freq_{it}$ before the change in the PMS. There was no impact of adding the worker satisfaction measure to the contract, i.e., β_{f2} was not significant ($t=1.57$). This is in contrast with the result from the yearly analysis (see section 6.6.2), where a positive impact was found. There are a number of possible explanations for this surprising result. First, $Freq_{it}$ was decreasing already before the change and since it is a ratio-measure it becomes increasingly more difficult to improve when the measure goes down²⁴. This is especially the case since the baseline, i.e., T_t , is assumed to be linear. Second, worker satisfaction had a negative impact on future costs (see section 5.4.4). Therefore, there is a trade-off between improving on worker satisfaction and costs, and the managers might have valued improvements in costs more.

Parameter β_{f4} of $Rzat_{it}$ is negative and significant, -7.675 ($t=-10.06$, $p<0.01$) indicating that workers on Saturday are more satisfied. Finally, the control for influenza epidemics (IAZ_{it}) is significant, β_{f5} is 0.307 ($t=16.69$, $p<0.01$), indicating that during epidemics workers are more often sick for a short period.

²³ Remember that $Freq_{it}$ is negatively related with worker satisfaction.

²⁴ The $Freq_{it}$ measure decreased with $0.006 \cdot 36 = 0.216$ for the period under study, whereas the average of the measure was 1.36.

β_{o2} of equation (6.16) measures the impact of the change in the PMS on on-time delivery. The on-time delivery measure OTD_{it} has a positive trend before the change in the PMS system, represented by a β_{o1} of 0.001 ($t=4.84$, $p<0.01$). The change in the PMS had a positive impact on the on-time delivery measure, represented by a significant β_{o2} of 0.044 ($t=7.04$, $p<0.01$).

In sum, the data suggest a mixed impact of the change in the PMS on non-financial performance. There was no impact on the added worker satisfaction measure, the on-time delivery measure however increased after the change in the PMS.

6.6.4 Impact of change in PMS on financial performance

Table 6.6 reports the results of the impact of the change in the PMS on the financial measures.

Table 6.6: Impact of the change in PMS on financial performance

Variables ¹	<i>Costs_{it}</i> (n=972 ²)			<i>Rev_{it}</i> (n=972 ²)		
	Expected sign	Coefficient		Expected sign	Coefficient ^{3,4}	
Impact variables:						
D_t	?	-2.892 (-9.31)	**	+	0.426 (5.23)	**
Controls:						
T_t	?	0.037 (2.46)	**	?	0.027 (6.92)	**
Vol_{it}	+	0.001 (22.87)	**			
R ² -adj.		0.980			0.984	
DW-statistic ⁵		1.75			2.14	

Notes:

¹ The month dummies and intercepts for each area are considered to be nuisance variables, therefore they are not reported.

² 3 years of monthly observations of 27 areas.

³ Where **, * means significant at the 0.01, and 0.05 levels (two-tailed test).

⁴ t-statistics between parentheses. All results are from the seemingly unrelated regression model.

⁵ Where the DW-statistic means the Durbin-Watson statistic.

β_{c2} of equation (6.17) measures the impact of the change in the PMS on costs. There was a positive trend in the cost measure before the change in the PMS, expressed by a significant β_{c1} ($t=2.46$, $p<0.01$). The change in the PMS had a positive impact on costs, that is β_{c2} is negative (-2.892) and significant ($t=-9.31$, $p<0.01$). Because the production volume is almost given for the area managers (see section 3.6), I controlled for the number of products, Vol_{it} , in each time period. This parameter was positive, 0.001 ($t=22.87$, $p<0.01$), as expected.

β_{r2} of equation (6.18) measures the impact of the change in the PMS on revenues. The trend in Rev_{it} before the change in the PMS was significant, i.e., 0.027 ($t=6.92$, $p<0.01$). The change seems to have a positive impact on revenues expressed by a significant parameter β_{r2} of 0.426 ($t=5.23$, $p<0.01$).

In sum, the data suggest that the change in the PMS has a positive impact on both costs and revenues.

6.6.5 Managerial significance

Since the number of observations is large, even very small effects become significant (Rosenthal and DiMatteo, 2001). Therefore, the managerial significance of the results is expressed by computing the improvement of the change in the PMS as a percentage of the dependent variable, i.e., the coefficient that estimated the impact of the change divided by the average of the dependent variable. Results for this analysis are reported in table 6.7.

Table 6.7: Managerial significance of the change in the PMS

Variable	Impact parameter (D_t)	Managerial significance (%) ¹
<i>Freq_{it}</i>	0	0
<i>OTD_{it}</i>	0.044	4.9
<i>Cost_{it}</i>	-2.892	-5.2
<i>Rev_{it}</i>	0.426	1.4

Notes:

¹ The managerial significance of the change in PMS is measured as a % of the average dependent variable, i.e., this is the parameter that estimates the impact divided by the average dependent variable multiplied by 100%.

The change in the PMS led to a 4.9% increase in the on-time delivery measure. Considering the fact that this measure was high already, i.e., the average is 90%, and the benchmark against which the change in the PMS is compared, i.e., T_t , is linear²⁵, this can be considered a high value. The cost measure decreased with -5.2% of the average dependent variable after the change in the PMS. The impact of the change in the PMS on the revenues was an increase of only 1.4% of the average revenues.

The managerial impact of the change in the PMS from this study is comparable with Banker et al. (2000b). In their study they find improvements of 1.1% and 9.8% of the average dependent variables for the two non-financial measures that are indicators for customer satisfaction. The impact on the costs and revenues is -1.68% and 1.78%.

6.7 Summary and conclusion

This chapter assesses the impact of a change in the performance measurement system on managerial performance. Before the change the main emphasis in the contracts of middle managers was on financial measures, after the change the financial measures were added with a substantial emphasis on non-financial performance measures.

The change in the PMS had a positive effect on the financial measures costs, and revenues, i.e., cost decreased and revenues increased. However, of the non-financial measures there was only a positive impact on on-time delivery. The change in the PMS had no impact on the proxy used for the worker satisfaction, i.e., *Freq_{it}*. This is surprising since worker satisfaction was the measure that was added to the contract of the manager. There are a number of explanations for this result. First, the results from

²⁵ This means that the baseline improved with $0.001 \cdot 36 \cdot 100\% = 3.6\%$ already. The impact of 4.9% of the change in the PMS is additional to this improvement.

the incremental information content analysis (see section 5.4.4) show that worker satisfaction had a negative impact on future costs. Therefore, the managers have to make a tradeoff between improving costs and worker satisfaction. Second, the worker satisfaction measure improved already before the change in the PMS and therefore it was difficult for the managers to improve it further. Third, it is possible that worker satisfaction only had a positive impact via the other measures. To assess such indirect effects, path analysis should be used. Finally, the frequency index may not have been a good proxy for worker satisfaction. The impact of the change in the PMS on the measures was a 5% improvement for the cost, and on-time delivery measure, and a 2% improvement for the revenues.

There are a number of limitations to this research. First, although a number of variables and statistical techniques are used to control for threats to internal validity, a limitation is not having a natural control group to estimate the performance impact. This is a problem of all field studies where the change under study was not under control of the researcher. Second, the proxy used for the variable, for which no monthly time-series were available (i.e., absence-frequency for worker satisfaction), was identified from the literature. Although the correlation between both variables in the sample was higher than the effect-size that meta-studies indicate, the quality of the proxy can be discussed. Third, the company changed the PMS to align the contracts of the managers with the increased quality orientation in the organization. Therefore the problem of endogeneity exists, that is both are choice variables for the organization (Ittner and Larcker, 2001).

Despite these limitations, this paper shows that many insights can be gained to observe effects of changes in management accounting practices. A number of directions for further research could be considered. First, results of this study should be corroborated in different settings. Ittner and Larcker (1997) give as a further research stream the evaluation of BSC-type of performance measurement systems. However, since it is almost impossible to evaluate changes in the PMS with comparable time-series data it is difficult to collect a large sample. Generalization of result from different studies is therefore probably only possible through theoretical generalization (Yin, 1989). Further, research could consider such efforts. Second, the PMS system in this company was used for both decision making and decision control purposes (Zimmerman, 1997). However, both purposes have different demands. Research that considers these different demands, and ascertains the problems with these dual purposes of one PMS might be helpful.

Although the results from the studies are specific for this case, the emphasis on the quality strategy through the whole organization, operationalized by the EFQM-model, is an important moderator for the relationship between non-financial performance measures use and performance (Ittner and Larcker, 1995, 1997).

Chapter 7

Summary and conclusion

7.1 Introduction

This chapter summarizes the content of the dissertation. The next section gives a brief overview of each chapter. Section 7.3 states the conclusions from the empirical studies and its implications. Section 7.4 enumerates the contributions of the research. Section 7.5 addresses the limitations of the research. Section 7.6 gives the directions for further research that can be derived from this dissertation.

7.2 Summary

This dissertation addresses two general research questions. First,

“Do non-financial performance measures have relative or incremental information content, or both, beyond financial performance measures in predicting future financial performance?”

Second,

“Do managers perform better when non-financial measures are added to their performance evaluation system?”

The first question is motivated through an existing ambiguity in the literature. Although it is often stated that non-financial measures are better indicators for future financial performance, the empirical literature always analyzes the incremental predictability of the non-financial measures beyond the financial measures.

The second research question is motivated by a direction of further research given by Foster and Young (1997). They argued that a lot of innovations are proposed in management accounting literature, but that a systematic evaluation of these innovations is mostly absent. This is in contrast with other scientific disciplines in which innovations are not implemented without a thorough evaluation. The second research question evaluates such an evaluation.

The empirical analysis took place in a large Dutch logistic company that had a substantial change in the PMS. Before the change managers were evaluated on financial measures, whereas after the change a substantial part of the evaluation was based on added non-financial measure. The data are from 27 geographically dispersed areas of the company. Three years of monthly data is used.

Chapter 2 gives a broad review of the available literature about the two research questions. In this chapter I review studies that discuss relationships between financial and non-financial performance measures, and studies that explore the use and usefulness of non-financial performance measures. From this review I conclude first that most research that explores relationships between financial and non-financial performance measures find empirical support for such relationships. However, these results are found in studies that are hardly comparable with each other, since each study uses a different definition for the lag, uses different specification, e.g., lag versus level models, and the results are found in different samples.

Second, I argue that the inconclusive results from performance effects of emphasizing non-financial measures relatively more might be explained by a number of limitations of the cross-sectional methodology. These limitations are the measurement of performance at one period and the short-term performance measures used. By adopting a longitudinal research methodology these limitations might be mitigated.

Chapter 3 gives background information about the company that might facilitate the understanding of the empirical analysis. First I describe the organization. Then, I describe the different elements of the PMS in the company, and explain the exact change in the PMS that took place in 1996. Finally, the controllability of the performance measures from the manager's contract is assessed.

Chapter 4 defines the performance data used in the empirical analysis. In addition, some properties of pooled time-series data, i.e., the data set used in the empirical analysis, are discussed. The biggest advantage of pooled time-series data is that it has more mechanisms to mitigate problems of heterogeneity. Finally, the descriptives of the data are reported.

Chapter 5 reports the empirical study that addresses the first research question. In this study, I test both the relative information content of non-financial measures compared with the financial measures, and the incremental information content of non-financial measures beyond financial measures. The relative information content question is the first test of the claim advocated in the popular management accounting literature that non-financial measures are better indicators for further financial performance measures than financial measures. The incremental information content analysis adds to the growing literature that assesses whether the non-financial measures explain future financial performance beyond the lagged financial measure. I suggest that the relative information content analysis is an important question to assess the relative weight of the different measures in the performance evaluation system. In contrast, results from the incremental information content question might be an input to answer the question, which measures need to get a weight in the contract.

Chapter 6 reports the empirical study that addresses the second research question. In this chapter a quasi-experimental study is executed that assesses the impact of the change in the PMS. Before the change managers were evaluated mainly on financial measures. After the change non-financial measures were included in the PMS.

7.3 Conclusions and implications

7.3.1 Information content of performance measures

In the first empirical study reported in chapter 5, I find that non-financial measures are not better predictors than lagged financial measures for future financial performance. In contrast, for the change models the lagged financial measures are better predictors. In addition, the expectation that the non-financial measure explains more when the number of lags increase is not supported by the results. Together, this implies that the claim that non-financial measures are better indicators for future financial performance than lagged financial measures is not true in this setting.

These results are summarized in table 7.1.

Table 7.1: Summary of results for the relative information content analysis

Information content of non-financial (IC_{nf}) versus financial (IC_f) measures for predicting future financial measures	Expectations	Results ¹
Future costs	$IC_{nf} > IC_{cost}$	n.s.
Future revenues	$IC_{nf} > IC_{cost}$	n.s.

Notes:

¹ n.s means that there was no significant difference in the information content between the financial and non-financial measures.

The non-financial measures worker satisfaction and on-time delivery measures have incremental information content beyond the lagged financial measures for future financial performance. More specific, a higher worker satisfaction measure leads to more future costs, but also to higher future revenues. A partial explanation for the result for future costs is that improving worker satisfaction costs more due to, for example, training, than it brings in the long-run. An explanation for the result for future revenues is that more satisfied workers provide better service to customers and therefore lead to more future revenues. Higher on-time delivery leads to lower future costs. This might be explained by the fact that when products are delivered at the wrong place, rework to get the product to the right place is costly. The on-time delivery measure also has a positive impact on future revenues. This again suggests that better service provided to the customer through a higher on-time delivery leads to more future revenues.

These results are summarized in table 7.2.

Table 7.2: Summary of result for incremental information content analysis

Non-financial measure → financial measure	Expectation ¹	Result ²
Worker satisfaction → future costs	?	+
On-time delivery → future costs	?	+
Worker satisfaction → future revenues	+	+
On-time delivery → future revenues	+	+

Notes:

¹ A + means that the non-financial measure is expecting to have a positive impact on future financial performance after controlling for the lagged financial performance. A ? means that the non-financial measure is expecting to have a differential impact on future financial performance after controlling for the lagged financial performance.

² A + means that the non-financial measure has a positive impact on future financial performance after controlling for lagged financial performance. A - means the non-financial measure has a negative impact on financial performance after controlling for future financial performance.

The non-financial measures give a substantial amount of incremental information content beyond the lagged financial measures. The incremental information content of the non-financial measures is roughly one-seventh for both worker satisfaction and on-time delivery compared with the information content of the lagged costs. For the revenues, the incremental information content is roughly one-fifth for each non-financial measure of the lagged revenues.

I also find that the incremental information content of non-financial measures is non-linear. For future costs, this means that when the level of non-financial measures is already high, an increase in non-financial measures leads to a lower change in the future financial measures than when the level of the non-financial measure is low. This means that it becomes increasingly more difficult, i.e., in terms of cost, to improve the non-financial measures when they are high already. An implication from this result is that there is an optimal level of the non-financial measure and that above this level cost increase. Although, the incremental information content for future revenues is also non-linear, it is opposite to the expected direction.

The results from the first empirical studies indicate that the non-financial measures are not better indicators than financial measures for future financial performance, but that the non-financial give substantial additional information beyond the financial measure. Theoretically, this implies that the non-financial measure should not get a higher weight in the contract than financial measures¹, but that they should get a non-zero weight.

In my opinion the results indicate that the almost mechanical statement that is often made that non-financial measures are better indicators for future financial performance than lagged financial measures should be reconsidered or at least used more cautiously.²

7.3.2 Evaluation of the change in the PMS

The expectations for the second empirical study are partly derived from the multi-task principal-agent model. However, this model uses the maximizing performance model, in which I use elements of the political and business model (see section 1.4). In addition, the principal-agent models assume that contracts are costless. Finally, the notion of the BSC assumes that non-financial measures are indicators of financial measures. This implies that adding non-financial measures in the contract of managers does not only have a direct effect on performance measures but might also have an indirect effect via non-financial measures on the financial measures. Together these elements lead to the result that it is difficult to derive uni-directional expectations.

The empirical results in chapter 6 show that the financial measures, i.e., costs and revenues, improve after the inclusion of non-financial measures in the managers' contract. In addition, the non-financial measure on-time delivery improved after the change. However, worker satisfaction, one of the measures that was added to the contract, did not improve after the inclusion in the contracts. There are a number of potential explanations for this surprising result. First, the empirical analysis in chapter 5 shows that the worker satisfaction measure had a negative impact on future costs. Thus, managers have to make a trade-off between improving costs or worker

¹ Again, this result only holds when measures have the same level of noise and sensitivity.

² This is an illustration of Luft and Shields (forthcoming) critique on using "practice defined variables". Without describing the underlying theoretical properties, non-financial measures are attributed a number of characteristics that are used to motivate its advantages.

satisfaction. Second, worker satisfaction improved already before the change in the PMS and therefore it is more difficult to improve it further after the change. Third, although I assess the association of the proxy used for worker satisfaction with yearly data, it is possible that the measure is not a proper proxy. Finally, it may be possible that there was no direct effect of the change in the PMS on worker satisfaction but that there was a beneficial indirect effect via worker satisfaction on on-time delivery and the financial measures.

These results are summarized in table 7.3.

Table 7.3: Summary of results of adding non-financial measures to the contracts of managers

Impact of change in PMS	Expectation¹	Result²
Non-financial measures ³		
Change in PMS → Worker satisfaction	+	n.s.
Change in PMS → On-time delivery	?	+
Financial measures		
Change in PMS → Costs	?	+
Change in PMS → Revenues	+	+

Notes:

¹ Where a + means that the change is expected to have a positive impact on the measure, a ? means that the change is expected to have a differential effect on the measure.

² Where a + means that the change has a positive impact on the measure, and n.s. means that the change had not a significant effect on the measure.

³ Only the worker satisfaction measure was added to the contract.

The change in the PMS had a substantial impact on the performance measures. The on-time delivery measure rose with 4.9%, the cost measure declined with 5.2% and the revenues rose with 1.4% of the average dependent variable.³

The tenet of these results seems to indicate that the change in the PMS had a favorable impact on most measures.

7.4 Contributions of the dissertation

This section enumerates the contributions of the research to the literature. First, the first empirical study addresses the difference between relative and incremental information content in a management control context. Although earlier research often assumed that non-financial measures are better indicators of future performance, i.e., have more relative information, the empirical analyses always assessed non-financial measures' incremental contribution beyond lagged financial measures.

Second, I add to the growing literature that assesses incremental information content of non-financial measures. Contributions to this stream of literature are that, I use different non-financial measures, and assess non-linearities in these relationships. In addition, the empirical study finds that the lag between non-financial measures and financial measures is different for the two non-financial measures and the two financial measures. This result contrast earlier research that finds the same lag between each non-financial measure and each financial measure. Further, in the methodology used for the lag search procedure I relax the constraint that the impact of the non-financial measure on the financial measure is the same for each period.

³ This percentage is computed by the parameter of the change in the PMS, divided by the average dependent variable multiplied by 100%.

Third, this study is one of the few studies that assess long-term impact of including non-financial performance measures in the PMS. This method might overcome at least two limitations of earlier research. First, most studies that are conducted before used short-term performance measures, e.g., return-on-assets, to evaluate the use of non-financial performance measures. A second related problem is that these studies all use a cross-sectional methodology to assess the performance consequences. Such a methodology might understate the positive effects.

Finally, the two empirical studies are executed in a different context. In the study of Banker et al. (2000b), customer satisfaction was the most strategic issue in their research setting. In my setting, the company stresses a quality orientation. In addition, in the study of Banker et al. (2000b) the compensation element considered is the bonus of managers. In contrast, in my research setting the performance measures is the input for salary increase, bonus allocation, and future promotions.

The contributions given above are all important for management accounting research. However there are also some contributions for practice. Especially, the first study shows a number of methods to assess whether non-financial measures are helpful in a particular situation. Therefore although the generalizability of the results might be limited (see section 7.5), the methods used could be adopted more widely.

7.5 Limitations of the study

As with all empirical studies, the results of the studies should be interpreted in the context of its limitations. In the first empirical study, the usefulness of incorporating the non-financial measures in the contract of the area managers is assessed through the relative and incremental information content of the measures. However weight of a measure in the contract also depends on the noise and sensitivity of the measure (Banker and Datar, 1989). In addition, the results are based on the information content of the measures and therefore do not incorporate other potential advantages of non-financial measures, for example, that non-financial measures are better understood than financial measures by lower level workers.

The method to compute a lag between non-financial and financial performance measures assumes that the causality is from improved non-financial measures to improved financial measures. However, to improve non-financial measures the area managers probably need to invest first, for example in redesigning processes to improve on-time delivery. Therefore, bi-directional methods should be used to assess this impact (Luft and Shield, forthcoming). However this raises the difficulty that it is necessary to estimate two lags.

Finally, the results are found in a sample of one company. Therefore, interpretation of the results should be done with caution. For example, in this context a higher on-time delivery leads to less future costs. However, in other contexts this relationship could equally well be positive depending on the circumstances.

The evaluation study also has a number of additional limitations. The most important limitation is the absence of a control group of areas in which the change in the PMS did not take place. Due too this absence, alternative methods are used to compare the impact of the change in the PMS against a benchmark. These alternatives give room for additional alternative explanations for the results found. The problem of the lack of a natural control group is a general problem in management accounting evaluation research. Researchers do not often get a change to implement a management accounting innovation in a number of departments or business units whereas the others are used as a control group. Hence, this type of research should be

accomplished without a limited control group or performance impacts of new management accounting systems should only be considered in cross-sectional studies.

The results found are specific for this company. Therefore any generalizability of the results to other companies, situations or samples should be made cautiously. However, the organization implemented the new PMS to align the PMS with the increased emphasis on a quality orientation. This quality orientation is a moderator for non-financial performance measure use and performance (Ittner and Larcker, 1995, 1997).

The fact that the change in the PMS was motivated by the increase quality orientation suggest that the study is subject to the endogeneity problem (Ittner and Larcker, 2001). This means that both the change in the PMS and increased quality orientation are both choice variables. Although the quality orientation is not explicitly used as a variable in the analysis, endogeneity might still have an impact on the results.

A limitation of the data is the absence of monthly observations for the two added performance dimensions. For one of the performance dimension, i.e., worker satisfaction, I used a proxy. Although the quality of this proxy is estimated, its validity can be questioned.

Finally, Cook and Campell (1979) suggest a number of potential effects that can be expected in an interrupted time-series design. Since there is no theory that can guide the expected effects on performance after the change in the PMS, for econometric reasons I ascertain these effects by modeling only a structural shift in the time-series in the specifications. However, this specification is clearly arguable.

7.6 Further research

The two empirical studies suggest a number of interesting research questions. First, the study in chapter 5 finds different lags between the 2 non-financial measures and the financial measures. It is unclear what factors explain these differences. Further research could consider what factors drive these differences in lags. Some variables that could be considered are the competitiveness of the industry, the type of products sold, etc. In addition, I find that the worker satisfaction measure has a negative impact on future costs, whereas the on-time delivery measure has a positive impact on future cost after controlling for the current costs. Further research could consider which types of measures have a negative or positive impact, or which characteristics of measures drive this negative or positive impact.

The relative information content analysis can be used as an input to compare the relative weight of different measures in the PMS. The incremental information content can be used as an input to the question which measures should get a weight in the PMS⁴. Further research could consider whether this distinction is really used in practice for the choice to select performance measures in contracts.

Second, the empirical accounting literature uses the change models often as a robustness check for level models. However both models test a different relationship, and therefore test different theories (see Lambert, 1998). Further research should consider these differences and use this distinction in theory building.

Also from the second empirical study some directions for further research come to mind. The PMS in the company that is studied had both decision making and decision control aspects. However, the literature review already indicated that these two

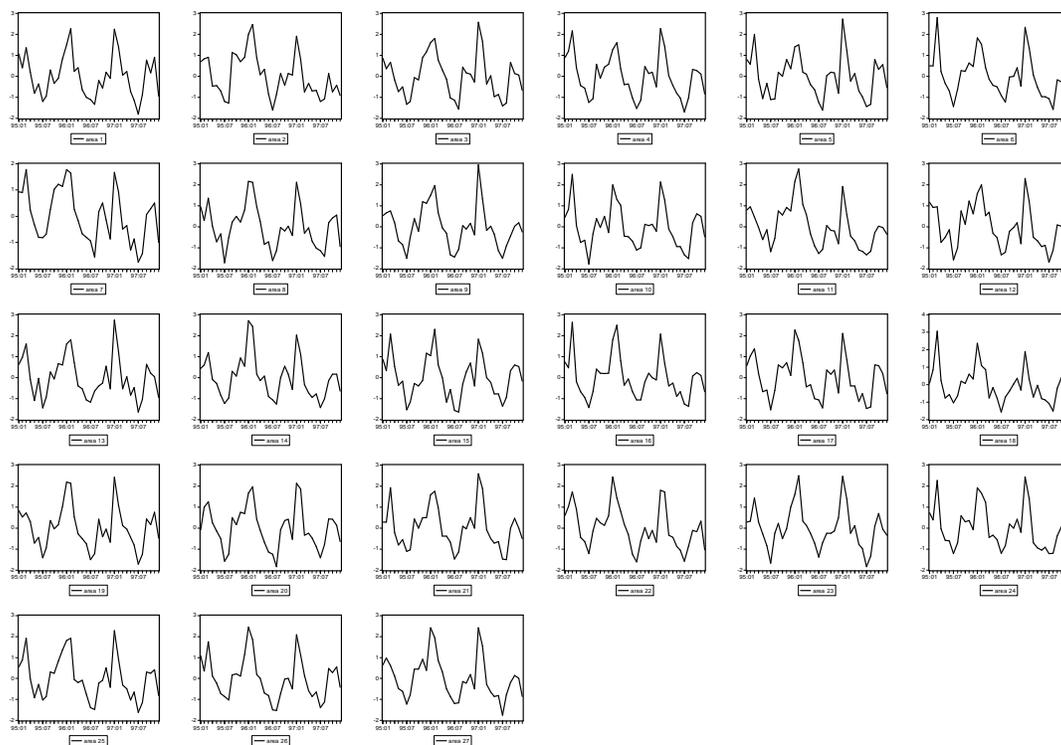
⁴ Again, this conjecture only holds when measures are noiseless and agents are risk-averse.

purposes of PMS systems can have a contrary impact. Therefore future research should acknowledge this difference and take it into consideration in empirical work.

To my knowledge there are two longitudinal studies that assess the performance impact of including non-financial performance measures into the contract of managers in the literature, i.e., Banker et al. (2000b) and this study. Both studies use different samples and are accomplished in a different context. Therefore, the generalizability is limited in both studies and efforts to theoretically generalize the results might be helpful. In general, this stream of research is not characterized by strong theoretical underpinnings, this manifests itself, for example, by the high number of differential expectations. Efforts to strengthen this body of theory, if possible, would be most helpful.

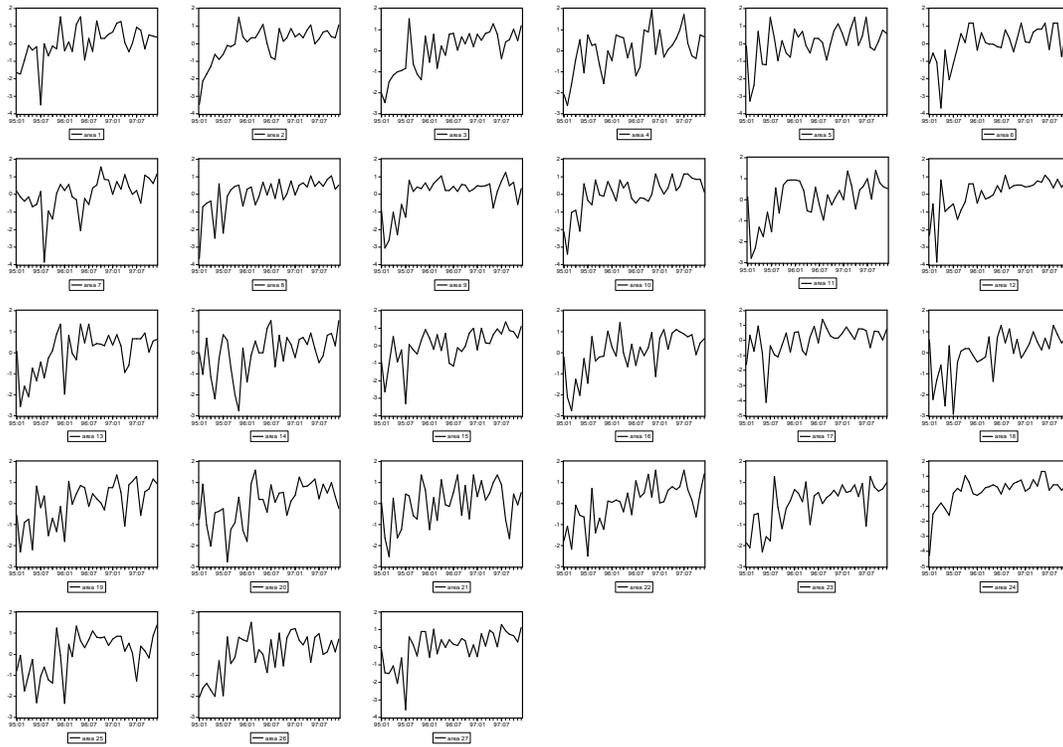
Appendix: Graphs of data

Figure A.1 Standardized frequency-index of each area



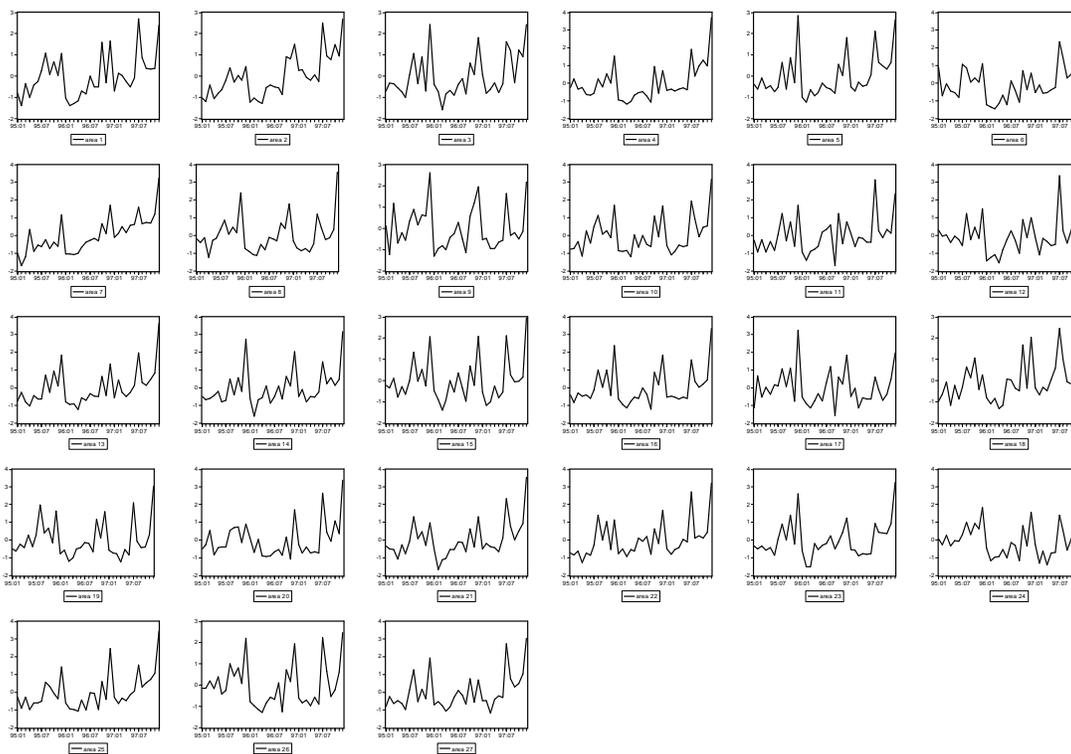
Notes:

These are the standardized observations of $Freq_{it}$ of the 27 areas. The y-axis is the number of standard deviations from the area mean. The x-axis is the time period.

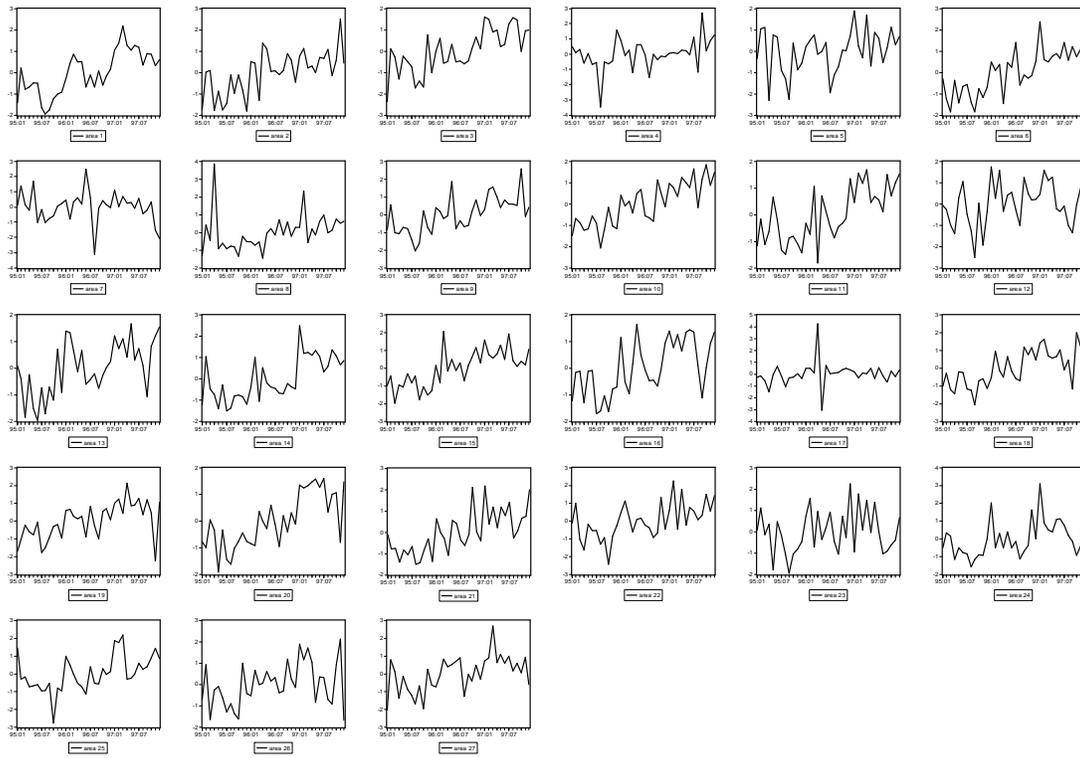
Figure A.2 Standardized on-time delivery of each area**Notes:**

These are the standardized observations of OTD_{it} of the 27 areas. The y-axis is the number of standard deviations from the mean. The x-axis is the time period.

Figure A.3 Standardized costs of each area

*Notes:*

These are the standardized observations of $Cost_{it}$ of the 27 areas. The y-axis is the number of standard deviations from the area mean. The x-axis is the time period.

Figure A.4 Standardized revenues of each area.**Notes:**

These are the standardized observations of Rev_{it} of the 27 areas. The y-axis is the number of standard deviations from the area mean. The x-axis is the time period.

References

Abernethy, M. A. and A. M. Lillis, (1995), The impact of manufacturing flexibility on management control systems design, *Accounting, Organizations and Society*, Vol. 20, p. 241-258.

Akaike, H., (1973), *Information Theory and an Extension of the Maximum Likelihood Principle*, Second international symposium on information theory, Budapest.

Allison, P. D., (1990), Change Scores as Dependent Variables in Regression analysis, in: *Sociological Methodology*, Washington, DC, American Sociological Association.

Amir, E. and B. Lev, (1996), Value-relevance of non-financial information: the wireless communications industry, *Journal of Accounting and Economics*, Vol. 22, p. 3-30.

Anderson, E., C. Fornell and D. Lehmann, (1994), Customer satisfaction, market share, and profitability, *Journal of Marketing*, July, p. 53-66.

Anderson, S. W. and S. M. Young, (1999), The impact of contextual and process factors on the evaluation of activity-based costing systems, *Accounting, Organizations and Society*, Vol. 24, p. 525-559.

American Accounting Association, (1975), Report of the committee on non-financial measures of effectiveness, *American Accounting Association*, Sarasota, Flo.

Baiman, S., (1990), Agency research in managerial accounting: a second look, *Accounting, Organizations and Society*, Vol. 15, p. 341-371.

Baker, G. P., M. C. Jensen and K. J. Murphy, (1988), Compensation and incentives: practices vs. theory, *Journal of Finance*, Vol. 63, p. 593-616.

- Baltagi, B. H., (1995), *Econometric Analysis of Panel Data*, New York, NY, John Wiley and Sons.
- Banker, R. D. and S. M. Datar, (1989), Sensitivity, precision, and linearity aggregation of signals for performance evaluation, *Journal of Accounting Research*, Vol. 27, p. 21-39.
- Banker, R. D., J. M. Field, R. G. Schroeder and K. K. Sinha, (1996a), Impact of work teams on manufacturing performance: a longitudinal study, *The Academy of Management Journal*, Vol. 39, p. 867-890.
- Banker, R. D., S. Lee and G. Potter, (1996b), A field study of the impact of a performance-based incentive plan, *Journal of Accounting and Economics*, Vol. 21, p. 195-226.
- Banker, R. D., S. Lee, G. Potter and D. Srinivasan, (2000a), An empirical analysis of continuing improvements following the implementation of a performance-based compensation plan, *Journal of Accounting and Economics*, Vol. 30, p. 315-350.
- Banker, R. D., S. N. Janakiraman and, C. Konstans, (2001), Balanced Scorecard: linking strategy to performance, *FEI Research Foundation*.
- Banker, R. D. and R. Mashruwala, (2002), Simultaneity in the Relationship between sales performance and components of customer satisfaction, *Working Paper*, University of Texas at Dallas, and Washington University at St. Louis.
- Banker, R. D., G. Potter and G. Schroeder, (1993), Reporting manufacturing performance measures to workers: an empirical study, *Journal of Management Accounting Research*, Vol. 5, p. 33-55.
- Banker, R. D., G. Potter and D. Srinivasan, (2000b), An empirical investigation of an incentive plan that includes nonfinancial performance measures, *The Accounting Review*, Vol. 75, p. 65-92.
- Banker, R. D. and A. Theveranjan, (2000), Goal congruence and evaluation of performance measures, *Working Paper*, University of Dallas, and Syracuse University.
- Beck, N., (2001), Time-series-cross-section data: what have we learned in the past few years, *Annual Review of Political Science*, Vol. 4, p. 271-293.
- Becker, B. and B. Gerhart, (1996), The impact of human resource management on organizational performance: progress and prospects, *The Academy of Management Journal*, Vol. 39, p. 779-801.
- Behn, B. K. and R. A. Riley, (1999), Using non-financial information to predict financial performance: The case of the U.S. airline industry, *Journal of Accounting, Auditing and Finance*, Vol. 14, p. 29-56.

- Biddle, G. C., G. S. Seow and A. F. Siegel, (1995), Relative versus incremental information content, *Contemporary Accounting Research*, Vol. 12, p. 1-23.
- Bowen, H. P. and M. F. Wiersema, (1999), Matching method to paradigm in strategy research: limitations of cross-sectional analysis and some methodological alternatives, *Strategic Management Journal*, Vol. 20, p. 625-636.
- Brancato, C. K., (1995), New performance measures-a research report, *Report number 1118-95-RR*, New York, NY: the conference board.
- Bruns, jr., W.J. and S. M. McKinnon, (1993), Information and managers: a field study, *Journal of Management Accounting Research*, Vol. 5, p. 84.
- Carr, S., Y. Mak and J. E. Needham, (1997), Differences in strategy, quality management practices and performance reporting between ISO accredited and non-ISO credited companies, *Management Accounting Research*, Vol. 8, p. 383-403.
- Chenhall, R. H., (1997), Reliance on manufacturing performance measures, total quality management and organizational performance, *Management Accounting Research*, Vol. 8, p. 187-206.
- Coates, J. B., E. W. Davis, C. R. Emmanuel, S. G. Longden and R. J. Stacey, (1992), Multinational companies performance measurement systems: international perspectives, *Management Accounting Research*, Vol. 3, p. 133-150.
- Cochrane, D. and G. H. Orcutt, (1949), Application of least-squares regressions to relationships containing autocorrelated error terms, *Journal of Statistical Association*, Vol. 44, p. 32-61.
- Cohen, J. and P. Cohen, (1983), *Applied Multiple Regression Correlation Analysis for the Behavioral Sciences*, Hillsdale, NJ, Lawrence Erlbaum Associates, Publishers.
- Cook, T. D. and D. T. Campbell, (1979), *Quasi-experimentation: Design & Analysis Issues for Field Settings*, Boston, Mass., Houghton Mifflin.
- Cook, T. D. and W. R. Shadish, (1994), Social experiments: some developments over the past fifteen years, *Annual Review of Psychology*, Vol. 45, p. 545-580.
- Datar, S., S., Cohen Kulp and R. A. Lambert, (2001), Balancing performance measures, *Journal of Accounting Research*, Vol. 39, p. 75-92.
- Dechow, P., (1994), Accounting earnings and cashflows as measures of firm performance: the role of accounting accruals, *Journal of Accounting and Economics*, Vol. 18, p. 3-42.
- Doran, H. E. and W. E. Griffiths, (1983), On the relative efficiency of estimators which include the initial observations in the estimation of seemingly unrelated regressions with first-order autoregressive disturbances, *Journal of Econometrics*, Vol. 23, p. 165-191.

- Eccles, R. and S. Mavrinac, (1995), Improving the corporate disclosure process, *Sloan Management Review*, Vol. 36, p. 11-25.
- Emsley, D., (2000), Variance analysis and performance: two empirical studies, *Accounting, Organizations and Society*, Vol. 25, p. 1-12.
- Feltham, G. A. and J. Xie, (1994), Performance measure congruity and diversity in multi-task principle/agent relations, *The Accounting Review*, Vol. 69, p. 429-453.
- Fisher, J., (1992), Use of non-financial measures, *Journal of Cost Management*, Spring, 31-38.
- Foster, G. and D. W. Swenson, (1997), Measuring the success of activity-based cost management and its determinants, *Journal of Management Accounting Research*, Vol. 9, p. 109-141.
- Foster, G. and S. M. Young, (1997), Frontiers of management accounting research, *Journal of Management Accounting Research*, Vol. 9, p. 63-77.
- Gjesdal, F., (1981), Accounting for stewardship, *Journal of Accounting Research*, Vol. 19, p. 208-231.
- Govindarajan, V. and A. K. Gupta, (1985), Linking controls systems to business strategy: impacts on performance, *Accounting, Organizations and Society*, Vol. 10, p. 51-66.
- Greene, W. H., (2000), *Econometric Analysis*, Upper Saddle River, NJ, Prentice Hall.
- Groot, T.L.C.M., and P. Smidt, (2000), PTT Post, in: *Cases in Management Accounting*, edited by T.L.C.M. Groot and K. Lukka, Essex, Pearson Education, p. 283-312.
- Hackett, R. D., (1989), Work attitudes and employee absenteeism: a synthesis of the literature, *Journal of Occupational Psychology*, Vol. 62, p. 235-248.
- Hair, J. F., R. E. Anderson, R. L. Tatham and W. C. Black, (1998), *Multivariate Data Analysis*, Upper Saddle River, NJ, Prentice-Hall International, Inc.
- Hartmann, F. G. H., (1997), Accounting for Performance Evaluation: Effects of Uncertainty on the Appropriateness of Accounting Performance Measures, *Dissertation*.
- Hartmann, F. G. H., (2000), The appropriateness of RAPM: towards the further development of theory, *Accounting, Organizations and Society*, Vol. 25, p. 451-482.
- Harvey, A. C. and J. Durbin, (1986), The effects of seat belt legalisation on British road casualties: a case study in structural time series modelling, *Journal of Royal Statistical Society A*, Vol. 149, p. 187-227.

- Hemmer, (1996), Accounting measures, *Journal of Management Accounting Research*, Vol. 8, p. 87-116.
- Holmstrom, B., (1979), Moral hazard and observability, *The Bell Journal of Economics*, Vol. 10, p. 74-91.
- Holmstrom, B. and P. Milgrom, (1991), Multitask principal-agent analyses: incentive contracts, asset ownership, and job design, *Journal of Law, Economics and Organizations*, Vol. 7, p. 24-52.
- Hopwood, A. G., (1972), An empirical study of the role of accounting data in performance evaluation, *Journal of Accounting Research*, Vol. 10, p. 156-182.
- Horngren, C. T., G. Foster and S. Datar, (1994), *Cost Accounting: A Managerial Emphasis*, Englewood Cliff, NJ, Prentice-Hall international, inc.
- Hsiao, C., (1990), *Analysis of Panel Data*, Cambridge, Cambridge university press.
- Huselid, M. A. and B. E. Becker, (1996), Methodological issues in cross-sectional and panel estimates of the human resource-firm performance link, *Industrial Relations*, Vol. 35, p. 400-422.
- Indjejikian, R. J., (1999), Performance evaluation and compensation research: an agency perspective, *Accounting Horizons*, Vol. 13, p. 147-157.
- Ittner, C. D. and D. F. Larcker, (1995), Total quality management and the choice of information and reward systems, *Journal of Accounting Research*, Vol. 33, p. 1-34.
- Ittner, C. D. and D. F. Larcker, (1997), Quality strategy, strategic controls, and organizational performance, *Accounting, Organizations and Society*, Vol. 22, p. 293-314.
- Ittner, C. D. and D. F. Larcker, (1998a), Are nonfinancial measures leading indicators of financial performance?: an analysis of customer satisfaction, *Journal of Accounting Research*, Vol. 36 (supplement), p. 1-35.
- Ittner, C. D. and D. F. Larcker, (1998b), Innovations in performance measurement: trends and research implications, *Journal of Management Accounting Research*, Vol. 10, p. 205-238.
- Ittner, C. D. and D. F. Larcker, (2001), Assessing empirical research in managerial accounting: a value-based management perspective, *Journal of Accounting and Economics*, Vol. 32, p. 349-410.
- Ittner, C. D., D. F. Larcker and M. V. Rajan, (1997), The choice of performance measures in annual bonus contracts, *The Accounting Review*, Vol. 72, p. 231-255.
- Johnson, H. T. and R. S. Kaplan, (1987), *Relevance Lost: the Rise and Fall of Management Accounting*, Boston, Mass., Harvard Business Press.

- Juran, J. M., (1979), *Quality Control Handbook*, New York, NY, McGraw-Hill.
- Kald, M. and Nilsson, F., (2000), Performance measurement at Nordic companies, *European Management Journal*, 18, p. 113-127.
- Kaplan, R. S. and D. P. Norton, (1992), The balanced scorecard - measures that drive performance, *Harvard Business Review*, Jan-Feb, p. 71-79.
- Kaplan, R. S. and D. P. Norton, (1993), Putting the Balanced Scorecard to work, *Harvard Business Review*, Sep-Oct, p. 134-147.
- Kaplan, R. S. and D. P. Norton, (1996), *The Balanced Scorecard: Translating Strategy Into Action*, Boston, Mass., Harvard Business School Press.
- Kaplan, R. S. and D. P. Norton, (2000), *The Strategy-Focused Organization*, Boston, Mass., Harvard Business School Press.
- Keasey, K. and R. Watson, (1987), Non-financial symptoms and the prediction of small company failure: a test of Argenti's Hypothesis, *Journal of Business Finance and Accounting*, Vol. 14, p. 335-354.
- Lambert, R. A., (1998), Customer satisfaction and future financial performance discussion of are nonfinancial measures leading indicators for financial performance? An analysis of customer satisfaction, *Journal of Accounting Research*, Vol. 26 (supplement), p. 37-46.
- Lambert, R. A., (2001), Contracting theory and accounting, *Journal of Accounting and Economics*, Vol. 32, p. 3-87.
- Lev, B. and R. Thiagarajan, (1993), Fundamental information analysis, *Journal of Accounting Research*, Vol. 31, p. 190-215.
- Luft, J. and M. D. Shields, (forthcoming), Mapping management accounting: making structural models from theory-based empirical research, *Accounting, Organizations and Society*.
- Luft, J. and M. D. Shields, (2000), The effects of financial and nonfinancial performance measures on judgement and decision performance, *Working paper*, Michigan State university.
- Malina, M. A. and F. H. Selto, (2001), Communicating and controlling strategy: an empirical study of the effectiveness of the Balanced Scorecard, *Journal of Management Accounting Research*, Vol. 13, p. 47-90.
- Marcantonio, R. J. and T. D. Cook (1994), Convincing quasi-experiments: the interrupted time series and regression-discontinuity design, *Handbook of Practical Program Evaluation*, San Francisco, Cal., Jossey-Bass

- McAdams, J. L. and E. J. Hawk, (1994), *Organizational Performance and Rewards: 663 Experiences in Making the Link*, New York, NY, American Compensation Association.
- McCleary, R. and R. A. Hay, (1980), *Applied Time Series Analysis for the Social Sciences*, Beverly Hills, Cal., Sage publications.
- McShane, S. L., (1984), Job satisfaction and absenteeism: a meta-analytic re-examination, *Canadian Journal of Administrative Sciences*, Vol. 1, p. 68-77.
- Merchant, K. A., (1998), *Modern Control Systems: Text and Cases*, New Jersey, NJ, Prentice Hall.
- Meyer, M. W. and V. Gupta, (1994), The performance paradox, *Research in Organizational Behavior*, Vol. 16, p. 309-369.
- Milgrom, P. and J. Roberts, (1992), *Economics, Organization & Management*, Englewood Cliffs, NJ, Prentice Hall.
- Milgrom, P. and J. Roberts, (1995), Complementarities and fit strategy structure and organizational change in manufacturing, *Journal of Accounting and Economics*, Vol. 19, p. 179-208.
- Miller, G. A., (1956), The magical number seven, plus or minus two: some limits on our capacity for processing information, *The Psychological Review*, Vol. 63, p. 81-96.
- Miles, R. E. and C. C. Snow, (1978), *Organizational Strategy, Structure and Process*, New York, NY, McGraw-Hill.
- Mohr, L. B., (1995), *Impact Analysis for Program Evaluation*, Thousand Oaks, Cal., Sage publications.
- Morisette, R., (1998), Financial and non-financial towards an integrative theory of information choices in organizations, *Working Paper*, University of Waterloo.
- Nagar, V. and M. V. Rajan, (2001), The revenue implications of financial and operational measures of product quality, *The Accounting Review*, Vol. 76, p. 495-513.
- Nilakant, V. and H. Rao, (1994), Agency theory and uncertainty in organizations: an evaluation, *Organization Studies*, Vol. 15, p. 649-672.
- Orwin, R. (1997), Twenty-one years old and counting, in: *Evaluation for the 21st Century: a Handbook*, p. Thousands Oaks, Cal., Sage publications.
- Otley, D. T., (1980), The contingency theory of management accounting: achievements and prognoses, *Accounting, Organizations and Society*, Vol. 5, p. 413-428.
- Otley, D. T. and A. J. Berry, (1980), Control organization and accounting, *Accounting, Organizations and Society*, Vol. 5, p. 231-246.

Parks, R. W., (1967), Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated, *Journal of the American Statistical Association*, Vol. 63, p. 500-509.

Pearce, J. L., W. B. Stevenson and J. L. Perry, (1985), Managerial compensation based on organizational performance: a time series analysis of the effects of merit pay, *Academy of Management Journal*, Vol. 28, p. 261-278.

Perera, S., G. Harrison and M. Poole, (1997), Customer focused manufacturing strategy and the use of operation-based non-financial performance measures, *Accounting, Organizations and Society*, Vol. 22, p. 557-572.

Pindyck, R. S. and D. L. Rubinfeld, (1998), *Econometric Models and Econometric Forecast*, Boston, Mass., McGraw-Hill.

Rosenthal, R. and M. R. DiMatteo, (2001), Meta-Analysis: recent developments in quantitative methods for literature reviews, *Annual Review of Psychology*, Vol. 52, p. 59-82.

Sayrs, L. W., (1989), *Pooled Time Series Analysis*, Thousands Oaks, Cal., Sage publications.

Schiff, A. D. and L. R. Hoffman, (1996), An exploration of the use of financial and non-financial measures of performance by executives in an organization, *Business Research in Accounting*, Vol. 8, p. 134-153.

Sim, K. L. and L. N. Killough, (1998), The performance effects of complementarities between manufacturing practices and management accounting systems, *Journal of Management Accounting Research*, Vol. 10, p. 325-346.

Simons, R., (1987), Accounting control systems and business systems: an empirical analysis, *Accounting, Organizations and Society*, Vol. 12, p. 357-374.

Singleton-Green, B., (1993), If it matters, measure it!, *Accountancy*, Vol. 111, May, p. 52-53.

Strivers, B. P., T. J. Covin, N. G. Hall and S. W. Smalt, (1998), How nonfinancial performance measures are used, *Management Accounting*, Feb, 44-49.

Symons, R. T. and R. A. Jacobs, (1995), A total quality management-based incentive system supporting total quality implementation, *Production and Operations Management*, Vol. 4, p. 228-241.

Van der Stede, W. A., C. W. Chow and T. W. Lin, (2001), Strategy, choice of performance measures, and performance, *Working Paper*, University of Southern California, and San Diego State University.

Venkatraman, N., (1989), The concept of fit in strategy research: towards verbal and statistical correspondence, *Academy of Management Review*, Vol. 14, p. 423-444.

- Vuong, Q. H., (1989), Likelihood ratio tests for model selection and non-nested model hypotheses, *Econometrica*, p. 307-333.
- Vroom, V. H., (1964), *Work and Motivation*, New York, NY, Wiley.
- White, H., (1980), A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, Vol. 48, p. 817-838.
- Winker, P., (2000), Optimized multivariate lag structure selection, *Computational Economics*, Vol. 16, p. 87-103.
- Wruck, K. H. and M. C. Jensen, (1994), Science, specific knowledge, and total quality management, *Journal of Accounting and Economics*, Vol. 18, p. 247-287.
- Yin, R. K., (1989), *Case Studies Research: Design and Methods*, Thousands Oaks, Cal., Sage publications.
- Zimmerman, J. L., (1997), *Accounting for Decision Making and Control*, New York, NY, McGraw-Hill.

Summary in Dutch

Dit proefschrift bestudeert aspecten van het nut van het gebruik van niet-financiële prestatemaatstaven in het prestatiemeting- en evaluatiesysteem van managers. Van niet-financiële maatstaven wordt onder andere beweerd dat ze meer naar lange termijn aspecten van beslissingen die managers nemen kijken, en meer aanknopingspunten voor oplossingen van potentiële afwijkingen ten opzichte van de planning geven. Alhoewel dergelijke beweringen vaak voorkomen in zowel wetenschappelijke- als vakliteratuur is er nog steeds relatief weinig onderzoek gedaan naar kenmerken en effecten van niet-financiële maatstaven.

De onderzoeksvragen die gesteld worden in dit proefschrift en de motivatie daarvan worden uiteengezet in hoofdstuk 1. De volgende onderzoeksvragen staan centraal. Ten eerste,

"Hebben niet-financiële prestatemaatstaven relatieve of additionele informatiewaarde, of beide, boven financiële maatstaven voor het voorspellen van toekomstige financiële prestaties?"

Ten tweede,

"Presteren managers beter wanneer niet-financiële prestatemaatstaven aan hun prestatiemeting- en evaluatiesysteem worden toegevoegd?"

De eerste onderzoeksvraag wordt gemotiveerd door een bestaande ambiguïteit in de literatuur. In de literatuur wordt vaak beweerd dat niet-financiële maatstaven betere voorspellers zijn voor toekomstige financiële prestaties dan financiële maatstaven. Dit suggereert dat niet-financiële maatstaven meer informatiewaarde hebben dan financiële maatstaven voor het voorspellen van toekomstige financiële prestaties. Empirische onderzoeken testen echter vaak de additionele informatiewaarde van niet-financiële prestatemaatstaven boven financiële maatstaven. Dat wil zeggen dat niet-financiële maatstaven informatiewaarde hebben voor het voorspellen van toekomstige financiële prestaties wanneer gecontroleerd wordt voor de huidige financiële maatstaven. Resultaten van dergelijke analyses

kunnen als startpunt dienen voor de gehanteerde weging van de verschillende maatstaven in de beoordeling van managers.

De tweede onderzoeksvraag komt tegemoet aan de veelvuldig gehoorde vraag naar meer evaluatie van management accounting innovaties. In andere wetenschappelijke disciplines worden nieuwe programma's, behandelingsmethodes, etc., voor gebruik grondig geëvalueerd op de effecten die ze hebben. In de management accounting literatuur worden effecten van "nieuwe" management accounting systemen echter zelden systematisch getest.

In hoofdstuk 2 wordt een uitgebreid literatuuroverzicht gegeven. Dit overzicht bespreekt studies die relaties tussen financiële en niet-financiële maatstaven onderzoeken, die onderzoeken welke factoren het gebruik van niet-financiële prestatie maatstaven stimuleren, en die beschrijven of prestatiemetingsystemen die de nadruk leggen op niet-financiële maatstaven succesvol zijn wanneer ze in de juiste omstandigheden worden gebruikt. Na het overzicht worden de sterke en zwakte punten van en tegenstrijdigheden in de verschillende studies besproken. Tenslotte geef ik in dit hoofdstuk enkele richtingen voor verder onderzoek aan.

Hoofdstuk 3 geeft achtergrondinformatie over de onderneming waar het onderzoek is uitgevoerd. In de onderneming vond een belangrijke verandering in het prestatiemetingsysteem (PMS) plaats. Voor de veranderingen werden managers voornamelijk op financiële prestatie maatstaven beoordeeld en beloond, terwijl na de veranderingen het PMS werd uitgebreid met een aantal niet-financiële prestatie maatstaven. Dit hoofdstuk beschrijft de onderneming, de analyse-eenheden die zijn bekeken, het prestatiemetingsysteem en veranderingen in dit systeem en het doel van de veranderingen.

De gebruikte data voor de empirische studies in hoofdstuk 5 en 6 worden beschreven in hoofdstuk 4. Aangezien de onderzoeksvragen gaan over veranderingen in de tijd is een gepoolde tijdreeks dataset verzameld. Van de 27 analyse-eenheden werden drie jaar maandelijkse maatstaven van zowel financiële als niet-financiële maatstaven verzameld. In dit hoofdstuk wordt verder nog aandacht besteed aan de unieke karakteristieken van een gepoolde tijdreeks dataset.

In hoofdstuk 5 wordt een antwoord gegeven op de eerste onderzoeksvraag. In dit hoofdstuk wordt bekeken of de twee niet-financiële maatstaven werknemerstevredenheid en on-time delivery voorspellende waarde hebben voor de toekomstige financiële prestaties op kosten en omzet. Eerst wordt bekeken welk type maatstaven, financiële of niet-financiële, meer voorspellende waarde heeft voor toekomstige financiële maatstaven. Vervolgens wordt geanalyseerd of niet-financiële maatstaven voorspellende waarde hebben voor toekomstige financiële maatstaven wanneer voor huidige financiële maatstaven is gecontroleerd.

De resultaten van deze studie tonen aan dat de niet-financiële maatstaven niet meer informatiewaarde hebben dan de financiële maatstaven om toekomstige financiële prestaties te voorspellen. Deze verwachting komt ook niet uit wanneer meerdere vertraagde periodes worden opgenomen in de modellen. Deze resultaten suggereren dat de vaak gehoorde claim dat niet-financiële maatstaven betere voorspellers zijn dan financiële maatstaven om toekomstige financiële prestaties te voorspellen in deze context niet opgaat.

De niet-financiële maatstaven hebben wel additionele informatiewaarde naast de financiële maatstaven om toekomstige financiële prestaties te voorspellen. Meer specifiek blijkt dat een verhoging van de werknemerstevredenheid leidt tot meer toekomstige kosten maar ook tot meer toekomstige omzetten. Een verhoogde on-time delivery blijkt zowel tot een verlaging van toekomstige kosten als een verhoging van

toekomstige opbrengsten te leiden. Voor deze test werd eerst een zogenaamde “lag search” procedure uitgevoerd. Deze procedure bekijkt tot hoever in de tijd de niet-financiële maatstaven voorspellende waarde hebben voor toekomstige financiële prestatie maatstaven. Deze procedure geeft aan dat de niet-financiële prestatie maatstaven werknemerstevredenheid en on-time delivery een verschillende “lag” hebben voor de kosten en opbrengsten. Verder is de relatie tussen niet-financiële prestatie maatstaven en toekomstige financiële prestaties niet-lineair. Voor zowel werknemerstevredenheid als on-time delivery is het moeilijker, c.q. leidt het tot een grotere stijging in de kosten, om deze maatstaven te verhogen wanneer ze al op een hoog niveau zitten. De resultaten voor toekomstige omzetten geven opmerkelijke resultaten weer. Hier leidt een verdere verhoging van zowel werknemerstevredenheid als on-time delivery die al op een hoog niveau zitten tot grotere verhogingen van toekomstige omzetten.

In hoofdstuk 6 wordt de tweede onderzoeksvraag beantwoord. In dit hoofdstuk wordt een quasi-experiment beschreven dat de verandering in het PMS zoals beschreven in hoofdstuk 3 evalueert. Deze evaluatie vindt plaats door de tijdreeksen van de verschillende maatstaven uit het contract van de managers voor de verandering te vergelijken met de tijdreeksen van de maatstaven na de verandering. Er zijn verschillende theorieën die kunnen motiveren waarom de toevoeging van niet-financiële prestatie maatstaven in het PMS tot beter presterende managers zou leiden. Ten eerste stelt het multi-task principaal-agentschap model dat de waarde van additionele prestatie maatstaven kan liggen in een grotere doelcongruentie van de manager en de eigenaar alsmede door de informatiewaarde die het heeft over oncontroleerbare effecten in de reeds gebruikte maatstaven. Verder geeft het idee van complementariteit aan dat opbrengsten van een bepaalde strategie groter zijn naarmate gelieerde systemen complementair zijn met de strategie. Aangezien de onderneming waar de empirische analyses plaats vinden een kwaliteit georiënteerde strategie nastreeft zal een dergelijke strategie meer opleveren wanneer gelieerde systemen zoals het PMS complementair zijn.

Deze studie geeft aan dat de financiële maatstaven, zowel kosten als omzetten, verbeteren nadat de niet-financiële maatstaven in de contracten van managers worden opgenomen. Verder verbeterde de niet-financiële maatstaf on-time delivery, die reeds voor de verandering in het PMS zat. De aan het PMS systeem toegevoegde maatstaf werknemerstevredenheid veranderde niet. Er zijn een aantal redenen aan te voeren voor dit opmerkelijke resultaat. Ten eerste bleek uit hoofdstuk 5 reeds dat een verhoging van werknemerstevredenheid tot meer toekomstige kosten leidde. Hieruit volgt dat managers dus een afweging moeten maken of ze de kosten willen verlagen of werknemerstevredenheid willen verhogen. Ten tweede steeg de werknemerstevredenheid reeds voordat het werd toegevoegd aan het PMS, dit maakt het moeilijk om de maatstaf nog verder te verbeteren. Tenslotte kan het zijn dat de verandering in het PMS geen direct effect had op werknemerstevredenheid maar wel een indirect positief effect via de werknemerstevredenheid op de on-time delivery maatstaf.

Hoofdstuk 7 vat de conclusies van het onderzoek samen en bespreekt de implicaties. Dit hoofdstuk wordt afgesloten met de beperkingen van het onderzoek en geeft richtingen aan voor verder onderzoek. Een van de belangrijkste beperkingen van het onderzoek is dat het is uitgevoerd binnen één onderneming. Theoretische generalisatie van de resultaten kan plaats vinden naar ondernemingen met een kwaliteit georiënteerde strategie.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

240. I.F.C. MULDER, Soil degradation in Benin: Farmers' perceptions and responses.
241. A.W. SALY, Corporate entrepreneurship. Antecedents and consequences of entrepreneurship in large established firms.
242. S. VAN VELZEN, Supplements to the economics of household behavior.
243. R.A. VAN DER GOOT, High performance linda using a class library.
244. E. KUIPER, The most valuable of all Capital. A gender reading of economic texts.
245. P. KLIJNSMIT, Voluntary corporate governance disclosures; An empirical investigation of UK practices.
246. P.J.G. TANG, Essays on economic growth and imperfect markets.
247. H. HOOGEVEEN, Risk and insurance in rural Zimbabwe.
248. A.J. VAN DER VLIST, Residential mobility and commuting.
249. G.E. BIJWAARD, Rank estimation of duration models.
250. A.B. BERKELAAR, Strategic asset allocation and asset pricing.
251. B.J. VAN PRAAG, Earnings management; Empirical evidence on value relevance and income smoothing.
252. E. PEEK, Discretion in financial reporting and properties of analysts' earnings forecasts.
253. N. JONKER, Job performance and career prospects of auditors.
254. M.J.G. BUN, Accurate statistical analysis in dynamic panel data models.
255. P.C. VERHOEF, Analyzing customer relationships: Linking relational constructs and marketing instruments to customer behavior.
256. C.S. BOS, Time varying parameter models for inflation and exchange rates.
257. A. HEYMA, Dynamic models of labour force retirement; An empirical analysis of early exit in the Netherlands.
258. S. DEZELAN, The impact of institutional investors on equity markets and their liquidity.
259. D.J. DEKKER, Network perspectives on tasks in account management.
260. F.G. VAN OORT, Agglomeration, economic growth and innovation. Spatial analyses of knowledge externalities in the Netherlands.
261. U. KOCK, Social benefits and the flow approach to the labor market.
262. D.J. BEZEMER, Structural change in Central European agriculture. Studies from the Czech and Slovak Republics.
263. D.P.J. BOTMAN, Globalization, heterogeneity, and imperfect information.
264. H.C. VAN DER BLONK, Changing the order, ordering the change. The evolution of MIS at the Dutch railways.
265. K. GERXHANI, The informal sector in transition. Tax evasion in an institutional vacuum.

266. R.A.J. BOSMAN, Emotions and economic behavior. An experimental investigation.
267. A.P. VAN VUUREN, Empirical analysis of job search using novel types of data.
268. H. VAN DE VELDEN, An experimental approach to expectation formation in dynamic economic systems.
269. L. MOERS, Institution, economic performance and transition.
270. N.K. BOOTS, Rare event simulation in models with heavy-tailed random variables.
271. P.J.M. MEERSMANS, Optimization of container handling systems.
272. J.G. VAN ROOIJEN, Flexibility in financial accounting; income strategies and earnings management in the Netherlands.
273. D. ARNOLDUS, Family, family firm, and strategy. Six Dutch family firms in the food industry 1880-1970.
274. J.-P.P.E.F. BOSELIE, Human resource management, work systems and performance: A theoretical-empirical approach.
275. V.A. KARAMYCHEV, Essays on adverse selection: A dynamic perspective.
276. A.J. MENKVELD, Fragmented markets: Trading and price discovery.
277. D. ZEROM GODEFAY, Nonparametric prediction: Some selected topics.
278. T. DE GRAAFF, Migration, ethnic minorities and network externalities.
279. A. ZORLU, Absorption of immigrants in European labour markets. The Netherlands, United Kingdom and Norway.
280. B. JACOBS, Public finance and human capital.
281. PH. CUMPERAYOT, International financial markets: Risk and extremes.
282. E.M. BAZSA-OLDENKAMP, Decision support for inventory systems with complete backlogging.
283. M.A.J. THEEBE, Housing market risks.
284. V. SADIRAJ, Essays on political and experimental economics.
285. J. LOEF, Incongruity between ads and consumer expectations of advertising.
286. J.J.J. JONKER, Target selection and optimal mail strategy in direct marketing.
287. S. CASERTA, Extreme values in auctions and risk analysis.
288. W.H. DAAL, A term structure model of interest rates and forward premia: An alternative approach.
289. H.K. CHAO, Representation and structure: The methodology of econometric models of consumption.
290. J. DALHUISEN, The economics of sustainable water use. Comparisons and lessons from urban areas.
291. P. DE BRUIN, Essays on modeling nonlinear time series.
292. J. ARDTS, All is well that begins well: A longitudinal study of organisational socialisation.
293. J.E.M. VAN NIEROP, Advanced choice models.
294. D.J. VAN VUUREN, The market for passenger transport by train. An empirical analysis.
295. A. FERRER CARBONELL, Quantitative analysis of well-being with economic applications.
296. L.M. VINHAS DE SOUZA, Beyond transition: Essays on the monetary integration of the accession countries in Eastern Europe.

297. J. LEVIN, Essays in the economics of education.