TI 2008-072/3
Tinbergen Institute Discussion Paper

The Degradation of Distorted Performance Measures: An Application to Residual Income Based Measures like *EVA*

Randolph Sloof
Mirjam van Praag

University of Amsterdam, and Tinbergen Institute.

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

The Degradation of Distorted Performance Measures:
An Application to Residual Income Based Measures like *EVA*

Randolph Sloof
Mirjam van Praag*

ABSTRACT: Baker (2002) has demonstrated theoretically that the quality of performance measures used in compensation contracts hinges on two characteristics: noise and distortion. These criteria, though, will only be useful in practice as long as the noise and distortion of a performance measure can be measured. Courty and Marschke (2007) have recently developed an elegant empirical test to detect distortion, based on the degradation of a performance measure subsequent to increasing its weight in the remuneration contract. We apply their test to assess the distortion of the often used class of performance measures that are based on 'Residual Income' (*RI*), such as 'Economic Value Added' (*EVA*). Residual income is widely used to measure and reward the performance of management boards. We use a difference-in-difference approach to account for (a) changes in economic circumstances in the period studied and (b) the self-selection of firms into the treatment and the control groups. Our results show that *RI* has degraded and is, therefore, a distortionary performance measure that can be gamed.

JEL-codes: D21, G35, J33, L21, M12, M41, M52

Keywords: Residual income, EVA, degradation, distortion, performance measurement, management compensation, incentive compensation

---

* This version: July 2008. Corresponding author: c.m.vanpraag@uva.nl; Roetersstraat 11, 1018 WB Amsterdam, the Netherlands. We acknowledge the research assistance by Dennis Schoenmakers and the comments by George Baker.

1. Introduction

Performance measurement is a crucial part of the design of any incentive system and is of strategic importance for most companies. Incentive systems may be an effective route to lower agency costs. Moreover, and most notably at the board level of publicly quoted companies, the required increased compliance to corporate governance codes and legislation have forced firms to tie pay to performance in increasingly transparent and measurable ways. This has put the measurement of the quality of performance measures high on the agenda of both practitioners and academics. This paper contributes to that item on the agenda.

The classic principal agent model acknowledges that the suitability of performance measures for use in compensation contracts depends on their noise. The noisier the performance measure, the lower is the optimal incentive intensity. However, as Baker (2002) points out, the critical issue in most incentive contracts may not be the noisiness of the performance measure involved, but rather its "distortion". Baker defines distortion inversely as the extent to which the effect of effort on the performance outcome is aligned with the effect of effort on the objective function of the firm. Distortionary performance measures do affect agents' behavior, but in an unintended direction, due to the possibility of increasing the level of the performance measure by other actions than the ones intended. These other actions are easier or cheaper from the agent's perspective, so that the activation of the performance measure in the contract will give incentives to put effort into these activities that increase measured performance but do not, or to a lesser extent, contribute to organizational value. In that case the performance measure has *degraded* due to its introduction into the reward function: The alignment between the measure and the value of the company has decreased. Distortion reflects the extent to which degradation may happen.

An example of the effect of employing a distortionary performance measure is the introduction of the performance measure "client satisfaction" in the reward function of a sales employee. Before this introduction, the sales rep's (perhaps suboptimal levels of) efforts to increase client satisfaction are likely to be aligned with the company's objective. As soon as the measure is used as a basis for performance pay, however, the sales rep will put (the least costly) effort into all actions that possibly increase the performance measure. One such 'cheap' action is selling at low prices or even giving products away for free. The performance measure "Client satisfaction" has degraded and turns out less useful than expected as a basis for performance pay.[1]

[1] Our theoretical understanding of distortion is mainly based on multi-tasking models, see Holmstrom and Milgrom (1991), Feltham and Xie (1994), Datar et al., (2001) and most recently Baker (2002).These models show an inverse relationship between the extent of distortion of the available performance measure and the efficient incentive intensity.

The distortion of a performance measure is, however, difficult to measure empirically and, so far, academics have not been of much help in developing empirical measures of distortion. Courty and Marschke (2007) are a recent exception. They have developed a general empirical test to detect distortion based on a simple theoretical framework. They apply this test to performance measures that have been introduced in the course of a natural experiment in a governmental program. They find weak support for the degradation of these specific measures upon introducing them into the reward function of agents who manage and execute the US Governmental Job Training Partnership Act.

In this paper we apply the empirical test as developed by Courty and Marschke (2007) to test for distortions in the performance measure *Residual Income (RI)* that many firms have introduced in the late nineties to measure and reward the performance of management boards. More specifically, we collected a sample of firms that have introduced the residual income based performance measure *EVA* as developed by Stern Stewart and Company. The copyrighted *EVA* doctrine prescribes that if *EVA* is used for evaluating and rewarding management, it should be the only measure used (Stewart, 1991). This limitation to *EVA* adopters warrants that one of the necessary conditions of the test developed by Courty and Marschke is satisfied, i.e., that at a specific point in time, the weight in the performance contract that is attached to *EVA* has increased relative to the weight attached to any other performance measure used. Unlike *RI*, *EVA* cannot be measured solely on the basis of firms' accounting data.[2] The measurement of *EVA* is based on *RI* but differs from the accounting based measure due to some –for researchers intransparent- discretionary and standard adjustments. Biddle et al. (1997) find that the correlation between *RI* and *EVA* is 0.90. We measure *RI* for the firms in our sample and refer to this as the accounting based value of *EVA*.

Contrary to the natural experiment application of Courty and Marschke, the introduction of *EVA* (or any other performance measure) is a deliberate strategic decision and not based on random assignment. Therefore, adapting the empirical test to address self-selection is required. Moreover, since *EVA* has been introduced by many firms in a particular period, viz. the late nineties, one should also account for possible changes in economic circumstances over the course of the period studied. For these two reasons we apply a difference-in-difference (DID) approach where we compare the outcomes before and after the adoption of *EVA* in the board's rewards function of the treatment group with suitable control groups consisting of firms that have not adopted *RI*-based measures. We find that (i) the decision to introduce *EVA* into the board's reward function is based on the correlation between the accounting based value of *EVA* and the assumed objective function of the company, i.e. *Relative Total Shareholder Returns* (*RTSR*), and (ii) that accounting based *EVA* is a distortionary performance measure that can be gamed.

The contribution of our study is that, as far as we know, we are the first to assess empirically the distortion of a performance measure that is (widely) used in companies. This assessment requires

The use of EVA data obtained from Stern Stewart is not possible in our application, see Section 3.

a DID approach to cope with changing economic circumstances and self-selection. The DID approach to the test by Courty and Marschke (2007) is of general applicability. Tests of distortion that are applicable to real world incentive contracts –such as those based on *RI*- are valuable: the distortion of a performance measure is claimed to be a highly relevant (inverse) criterion of its quality and the application of high quality performance measures is imperative for setting intense incentives, which in turn is an important source of work motivation (and therefore organizational value).

The remainder of this paper is structured as follows. The next section discusses the test of distortion developed by Courty and Marschke (CM) and our DID approach to it. In section 3 we introduce the widely used class of performance measures *RI* in general, and *EVA* in particular, and discuss our empirical measure of company value, i.e., *RTSR*. Section 4 discusses our sample of firms and their characteristics. Section 5 demonstrates the results from applying the CM test to assess the distortion of *RI*. Section 6 discusses the results and concludes.

2 Detecting the presence of distortion in a performance measure

The empirical tests CM derive to detect distortion require multiple observations of agent performance under different contracts. These multiple performance outcomes per contract are assumed to be generated in different production environments, such that the marginal productivity of effort varies across observations. Courty and Marschke mention various environments where these assumptions are valid. One of these pertains to managerial compensation in which the researcher observes accounting performance in a sample of firms that use the same accounting measure. Each firm is then treated as a different production environment. Another example they give is when performance is measured on a periodic basis – e.g. monthly in case of sales reps, or yearly in case of CEOs – and the marginal productivity of effort varies periodically. In that case each period can be taken as a different production environment. Our environment is a mixture of these two; we consider a panel of firms observed over a time span of various years.

To introduce the empirical test of distortion more formally, we closely follow Courty and Marschke and conceptualize each firm as a principal-agent relationship. The agent's actions influence two performance measures (p_i for $i=1,2$) and the value of the firm (V) in the following way:

$$p_{ijt}(e,g) = \nu_{0jt} \cdot e_{0jt} + (\nu_{ijt} + \eta_{ijt}) \cdot e_{ijt} + w_{ijt} \cdot g_{ijt} + \varepsilon_{ijt} \tag{2.1}$$

$$V_{jt}(e,g) = \nu_{0jt} \cdot e_{0jt} + \nu_{1jt} \cdot e_{1jt} + \nu_{2jt} \cdot e_{2jt} - (\zeta_{1jt} \cdot g_{1jt} + \zeta_{2jt} \cdot g_{2jt}) + \varepsilon_{Vjt} \tag{2.2}$$

where subscripts j and t refer to firm $j \in \{1,..,n\}$ at time $t \in \{t_0-k,..,t_0,..,t_0+k\}$.[3]

In the above equations e_{ijt} (where $i=0,1,2$) and g_{ijt} reflect the agent's productive and gaming efforts, respectively. Both are assumed to be costly, with effort costs equal to $\frac{1}{2}e_{ijt}^2$ and $\frac{1}{2}g_{ijt}^2$ for all i, j and t. With respect to the marginal productivities it is assumed that $\nu_{ijt} \geq 0, \nu_{ijt} + \eta_{ijt} \geq 0, w_{ijt} \geq 0, \zeta_{ijt} \geq 0$ and $E\eta_{ijt} = E\eta^3{}_{ijt} = 0$. Moreover, all these marginals are assumed to be uncorrelated. Finally, ε_{ijt} and ε_{Vjt} reflect additive noise terms (with zero mean), capturing influences on performance and firm value that are unrelated to the agent's efforts. These are also assumed to be orthogonal to the marginal productivities.

Given (2.1) and (2.2), e_{0jt} captures the dimension of effort common to the two performance measures and the firm's objective function V. Both e_{1jt} and e_{2jt} represent measure specific effort dimensions. These effort dimensions are imperfectly captured by the respective performance measures, because the marginal productivity of effort in the performance measures may differ from the marginal productivity of effort in the firm's objective function. For performance measure $i=1,2$ this happens when $\eta_{ijt} \neq 0$ for some j,t. In that case the performance measure is distorted, because effort is imperfectly valued at the margin. As argued by Courty and Marschke, the random marginal productivity captures the idea that agents may have incentives to exert good effort but not in the correct quantities.

A second, yet related channel of distortion occurs when $w_{ijt}>0$ for some i,j,t. In that case a *gaming* distortion arises. Gaming is defined as costly effort that increases the performance as measured (and rewarded) by the performance measure, but does not add at all to the firm's objective function. This is reflected by the fact that the gaming efforts g_{ijt} enter the value function V negatively, see equation (2.2). The empirical tests developed by CM are able to detect both types of distortions, but are not always able to distinguish between the two.

Clearly, the extent to which the performance measures (and thus the potential distortions therein) affect the agent's effort choices depends on the weights β_i (for $i=1,2$) of these measures in the agent's reward function. For given β_1 and β_2, Courty and Marschke derive the agent's optimal effort response, the performance outcomes for each measure p_{ijt} as well as the outcome of the firm's objective function V_{ijt}. The general test for distortion in a performance measure they propose is based on how the association between the latter two changes upon a change in the incentive weights β_i. Suppose therefore that at time t_0 the incentive weights are changed from (β_1^I, β_2^I) to $(\beta_1^{II}, \beta_2^{II})$. In particular, assume that at t_0 the principal introduces performance measure 2 (i.e. $\beta_2^{II} > \beta_2^I = 0$), instead of a different performance measure that was (possibly) used as a basis for remuneration before

[3] This state-contingent action model with linear production technologies is taken from Courty and Marschke (2007, Section 2). We have only adapted the notation to our application where a 'project' (labelled α in CM) corresponds to a firm j at a particular time t.

t_0 (i.e. $\beta_1^I \geq \beta_1^{II} = 0$). As a result, the relative weight β_2/β_1 on performance measure two increases. This corresponds to the situation we study in our empirical analysis, where sample firms introduced *EVA* (=p_2) as performance measure to replace their existing performance measure(s) (=p_1).

Courty and Marschke show that, if the statistical association between p_2 and V decreases upon an increase in the relative weight β_2/β_1, then p_2 is distorted. Here statistical association can either be measured by the (Pearson) correlation coefficient or by the slope coefficient of regressing V on p_2. If distortion is detected in this way, it can be of the two different kinds distinguished before. Thus both imperfect valuation of effort at the margin ($\eta_{2jt} \neq 0$ for some j,t) and explicit gaming distortions ($w_{2jt}>0$ for some j,t) are possible. But if the additional assumption is made that overall incentives are not weakened over time, then a decrease in the covariance between p_2 and V is conclusive evidence of gaming. Under the same assumption both the average value of p_2 and the variance in p_2 are predicted to increase. These last predictions can be used to establish that agents indeed respond to (changes in) incentives.

Let T=I denote the observed time span $\{t_0-k,..,t_0-1\}$ before the introduction of performance measure p_2 and T=II the observed time span $\{t_0+1,..,t_0+k\}$ afterwards. The relevant predictions CM derive can then be stated formally as follows:

Empirical CM tests. Suppose the relative weight β_2/β_1 increases when we move from T=I to T=II. Assuming that the joint distribution of the additive noise terms (ε_{ijt}, ε_{Vjt}) does not change with this change in performance weights,[4] it then holds that:

(Distortion):

(D.1) Measure 2 is distorted if $\Delta b = \left[\frac{Cov(p_{2jt}, V_{jt})}{Var(p_{2jt})}\right]_{T=II} - \left[\frac{Cov(p_{2jt}, V_{jt})}{Var(p_{2jt})}\right]_{T=I} < 0$;

(D.2) Measure 2 is distorted if $\Delta Corr = Corr(p_{2jt}, V_{jt})_{T=II} - Corr(p_{2jt}, V_{jt})_{T=I} < 0$;

(D.3) If $\beta_1^{II} + \beta_2^{II} \geq \beta_1^I + \beta_2^I$, then $\Delta Cov = Cov(p_{2jt}, V_{jt})_{T=II} - Cov(p_{2jt}, V_{jt})_{T=I} < 0$ implies the presence of a gaming distortion in measure 2.

(Incentives):

(I.1) If $\beta_1^{II} + \beta_2^{II} \geq \beta_1^I + \beta_2^I$, then $\Delta E(p_2) = E(p_{2jt})_{T=II} - E(p_{2jt})_{T=I} > 0$;

(I.2) If $\beta_1^{II} + \beta_2^{II} \geq \beta_1^I + \beta_2^I$, then $\Delta Var(p_2) = Var(p_{2jt})_{T=II} - Var(p_{2jt})_{T=I} > 0$.

[4] See Section 3.3 in Courty and Marschke (2007) where they make this assumption in their empirical model.

The CM tests for distortion thus require estimates of the level of association between p_2 and V, both before and after this performance measure has been introduced. The stated assumption on the invariance of the joint distribution of the additive noise terms when we move from T=I to T=II ensures that, in the theoretical model, changes in the level of association are solely guided by changes in the agent's effort choices. In practice, however, this may not be the case because for one reason or the other (see below) the invariance assumption may be violated. In that case, even when performance weights are unchanged, the estimated associations may change due to other factors when we move from T=I to T=II.

To address this issue in our empirical analysis, we compose suitable control groups to which the results of the treatment group of firms that did adopt *EVA* ($=p_2$) are compared. In particular, the treatment group consists of firms that have adopted *EVA* in period T=II as the (sole) performance measure that determines the board's remuneration, whereas it had not yet been in use before. To each firm in this treatment group we match a control firm that is 'similar' in a number of important respects, except that this control firm did not adopt an *RI*-based performance measure in the periods T=I and T=II studied.[5] Changes in the level of association observed for the control firms thus cannot be attributed to changes in the performance weights and must be due to other factors. Thus, by comparing the changes in association in the treatment group with the changes in association in the control group, we identify changes due to changes in effort incentives. We thus use a difference-in-differences approach (DID) to the empirical CM tests discussed above. For instance, to test (D.1) we look at whether $\Delta^2 b=[\Delta b]_{Treatment}-[\Delta b]_{Control}<0$. The other CM tests are adapted similarly.

Our first control group consists of firms that are similar to one of the treatment group firms in terms of size, industry and country (US).[6] For instance, treatment firm Coca Cola, which introduced *EVA* already in the early eighties, is matched to Pepsico Inc (see Table A1 in the Appendix for all pairs and their characteristics). By doing so we address the possibility that the association between p_2 and V and the first and second moments of p_2 may not only change after increasing its relative performance weight, but may also be affected by changes over time in economic circumstances and the like for the specific type of firm studied.

A second control group is composed to cope with self-selection. Using the overall sample of treatment and control firms as described above, a probit regression is estimated from which the observable determinants of the adoption of *EVA* are derived. Subsequently new treatment-control pairs are formed and matched based on the pre-treatment values of these variables (in line with propensity score matching).[7] There are two likely candidates that drive the adoption of a particular

[5] For each control firm the year t_0 that divides the two periods T=I and T=II equals the year in which the matched treatment firm adopted *EVA* as performance measure.
[6] Wallace (1997) uses the same set of criteria for selecting control firms. He studies the effect of introducing a residual income based performance measure on firm performance.
[7] Because control firms in this case are selected from the first control group, the second control group is a subset of the first.

performance measure, viz. noise and distortion. Noise is typically measured, by researchers and practitioners alike, as the (timeseries) variance in the performance measure due to factors beyond the agent's control (cf. Prendergast 2002, Coles et al., 2006). As noted by Baker (2002) and Courty and Marschke (2007),[8] distortion has so far wrongly been measured by the correlation between the value of the performance measure and the 'true' measure of the firm's objective (cf. Biddle et al., 1997; Stark and Thomas, 1998; Feltham et al., 2004). The higher this correlation is, the more likely the measure is deemed to be useful.

The most likely determinants of the use of a performance measure are thus intimately related to the statistical measures that CM propose to detect the distortiveness of the measure. If selection is based on the pre-treatment values of the variance of the performance measure and/or the correlation of this measure with the firm's objective, one would expect that the mean pre-treatment values of these statistics in the treatment group deviate from the mean levels over the same period in the firm population. In particular, if firms adopt performance measures that are highly correlated with the firm's objective in period T=I, as the evidence will show, one would expect, based on regression to the mean, a lower association between the performance measure and company value in period T=II for firms in the treatment sample. Therefore, decreasing correlation and regression coefficients are not necessarily indicators of distortion, but may as well reflect (self-selection induced) regression to the mean. The DID approach based on the second control group enables a distinction between regression to the mean and distortion.

3 *RI* based performance measures, *EVA*, and the firm's objective function *RTSR*

In this section we discuss the *RI* based performance measure *EVA* and our empirical measure of firm value, viz. relative total shareholder return (*RTSR*).

3.1 Residual Income (*RI*) and Economic Value Added (*EVA*)

Economists have long acknowledged that an appropriate measure of value creation for firms should be based on the difference between their earnings and the cost of capital (cf. Hamilton, 1777; Marshall, 1890; Biddle et al., 1997). In the twentieth century, various such concepts have been operationalized, including *residual income* (*RI*). Residual income is generally defined as net operating

[8] See e.g. Baker (2002, p. 736): "Consider first the large literature in accounting that has attempted to measure the correlation between various accounting measures and stock price. One objective of this research has been to determine whether incentive contracts should be based on accounting measures. […] measuring the correlation between accounting numbers and stock price is measuring the wrong thing. The correct measure (which is unfortunately much harder to assess) is whether accounting profits move with managers' actions in the same way that stock prices do." Zimmerman (1997) also criticizes the common practice of using a measure's correlation with stock returns as critical consideration in selecting an appropriate performance measure.

profits after-tax (*NOPAT*) minus a charge for invested capital (where operating profits are profits before deducting the after-tax costs of interest expenses). This capital charge equals the firm's weighted average cost of all debt and equity capital employed. Hence:

$$RI = NOPAT - WACC \cdot Capital\ Employed \tag{3.1}$$

Positive levels of *RI* thus reflect profits in excess of what is required by capital suppliers and implies value creation for the residual claimants of the company (i.e., the shareholders), whereas negative levels of *RI* are consistent with decreasing shareholder wealth.

More recently, in the 1980s, Stern Stewart & Company has developed and successfully advertized a trademarked version of residual income, labelled economic value added (*EVA*). Stern Stewart claims to have improved the measure by adjusting its elements, such that 'distortions' − like e.g. recording R&D as expenses rather than as capital investments − in the accounting model of performance measurement are taken away (Stewart, 1991, Chapter 2). Following the notation in Biddle et al. (1997, p. 306), *EVA* is operationalized as:

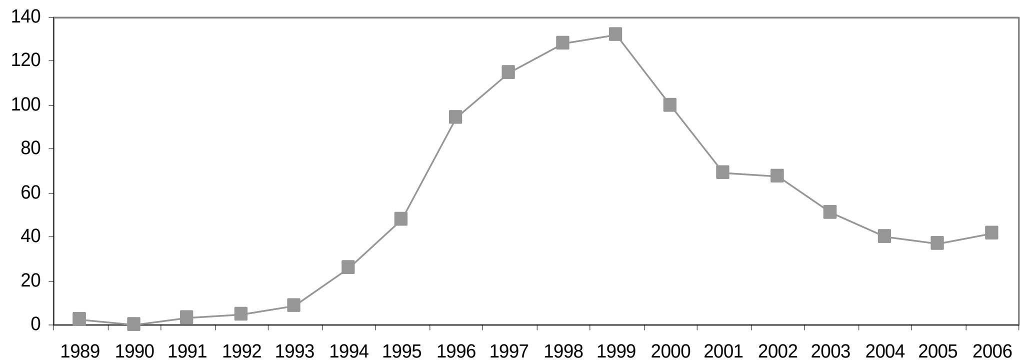
$$EVA = NOPAT + AccAdjOp - WACC \cdot [Capital\ Employed + AccAdjCa] \tag{3.2}$$

Here *AccAdjOp* reflects the Stern Stewart adjustments to accounting measures of operating profits and *AccAdjCa* their adjustments to accounting measures of capital employed.[9] Some of these adjustments are standard and others are discretionary.

Because the main goal of a for-profit organization is to add value to the owners' wealth in excess of their opportunity cost, *EVA* has bee promoted since the early 1990s as a key management tool. Many companies, such as Eli Lilly and Polaroid, started using *EVA* measures in their executive bonus schemes. Its popularity is indicated in Figure 1, which shows that the number of citations of *EVA* in the US business press grew exponentially in the mid 1990s. The LexisNexis database returns only two results in 1989 as compared to 132 in 1999. After the apogee of press citations in 1999, the number of press reports plunged. The later reports also seem to be much more critical in nature and some reports are concerned with scandals involving excessive bonuses. As we will discuss in Section 4, most of the companies in our treatment group introduced *EVA* in the mid 90s; the mean (median) year of introduction is 1995 (1996).

[9] Biddle et al (1997, p. 306) provide a large list of examples of such adjustments. Among others these include: the capitalization and amortization of research and development costs and of certain marketing costs, adding the change in bad debt allowances, adding the change in the LIFO reserve, adding (cumulative) goodwill amortization, and subtracting marketable securities and construction in progress.

Figure 1: Number of citation of "Economic Value Added" per annum in US press 1989-2006.



Stewart (1991) and Stern et al. (1995) posit that an *EVA*-based bonus plan should be based on incremental increases or the improvement in *EVA*, such that mere company size does not determine bonus levels. Remuneration committees use various methods to tie the executive compensation scheme to the improvement in *EVA* (*ΔEVA*). Some common practices are: (i) basing bonuses on the amount of excess *EVA* in a linear relationship, i.e. the nominal improvement of *EVA* from one year to the other, (ii) basing bonuses on the surplus realized above a minimum expected *ΔEVA* (of say 5 percent), or (iii) dictating ranges of *ΔEVA* and tie specific bonuses to them. Furthermore, although against the recommendations of the Stern & Stewart Company, many companies use upper limits to restrict bonus levels (typically to a maximum percentage of base salary). Finally, the copyrighted *EVA* doctrine prescribes that if *EVA* is used for evaluating and rewarding management, it should be the only measure used (Stewart, 1991). Section 4 will indeed show evidence that this recommendation has been widely followed in our sample of firms.

In our empirical analysis we perforce apply our tests to the accounting based value of *EVA*, i.e. the *RI* measure defined in equation (3.1). Annual *EVA* data covering the Russell 3000 US public companies are available from 1990 to 2007.[10] However, since most *EVA* adoptions took place in the mid nineties and our analysis requires sufficiently long timeseries of data both before and after the introduction of the performance measure, this dataset obtained from Stern Stewart and Company is not suitable. We recognize that the adjusted *EVA* in equation (3.2) may differ from the accounting based *EVA* calculation in equation (3.1). Biddle et al. (1997, p. 314) show that the correlation between the values of *EVA* based on the publicized numbers by Stern Stewart, i.e. only subject to standard adjustments, and the accounting based value of *EVA* amounts to 0.90.

[10] *EVA* in this dataset differs from *RI* only by the standard adjustments developed by Stern Stewart. As noted by Biddle et al., Stern Stewart makes additional custom adjustments for their corporate clients, which are not available to the public.

The three ingredients of accounting based *EVA* are *NOPAT*, *WACC* and *Capital Employed*. *NOPAT* is the operating profit which accrues to shareholders and debt holders and can be calculated in the following two ways:

$$NOPAT = Operating\ Profit \cdot (1 - Tax\ Rate) \quad (3.3a)$$

$$NOPAT = Operating\ Profit - Taxes\ paid\ (received) \quad (3.3b)$$

While in theory both approaches would logically yield the same result, in practice they are slightly distinct. For example, negative tax rates are not revealed in the Thompson WorldScope database that we use to calculate *NOPAT*, thus rendering the reimbursement of taxes impossible if using equation (3.3a) but not if using equation (3.3b). This study takes the average of the two.

The weighted average cost of capital (*WACC*) is defined as:

$$WACC = E/(E{+}D) \cdot r_e + D/(E{+}D) \cdot r_d \cdot (1 - T) \quad (3.4)$$

The *WACC* essentially weighs the required return on equity r_e and the required interest rate on debt r_d according to the relative weight of equity level E and debt level D in the financial structure of the company. Because debt interest payments are tax deductible, an adjustment for the tax rate T is enforced. The required rate of return on equity r_e can be determined using the capital asset pricing model:

$$r_e = r_f + \beta \cdot (r_m - r_f) \quad (3.5)$$

Here r_f refers to the risk free interest rate and r_m to the market rate. The extent to which the risk free interest rate should be increased to (or above) the market rate for a given company to arrive at the required return on equity, is determined by this company's market β.[11]

A couple of problems arise when calculating the *WACC* using annual financial data. Most importantly, judgmental decisions guide, for instance, the length and periodicity of the periods included in the calculation of β. Second, the required interest rate by debt holders is underestimated in case a company has suspended interest payments for one reason or the other (e.g., it may have obtained a head start from debt providers) and overestimated if a company has accelerated interest payments. Furthermore, various legitimate choices are at hand for the levels of the risk-free rate and the risk spread ($r_m - r_f$) that affect the calculated *WACC* levels. Altogether, equation (3.4) may lead to

[11] β is an indirect proxy for the level of non-diversifiable risk incorporated in a company's projects and directly captures the extent to which a company's stock returns are affected by changes in macroeconomic conditions. The higher β, the higher is the required rate of return on equity. For a firm with a market β higher than 1 (below 1), the total stock return will react more strongly (less strongly) than the market to macro-economic changes. The effect on a firm's total stock return will be equal to the average firm in the market if $\beta = 1$.

calculated *WACC* levels as low as 2.9%, which undoubtedly is a vast underestimation, and as high as 93.4%, which is exceptionally high to say the least.[12] Acknowledging these problems, remuneration committees of virtually all companies select predetermined or 'pre-calculated' levels for their *WACC*. We follow their practice and adopt a flat *WACC* for all companies in our study of 8, 10, 12, 14, and 16 percent, thereby covering the various *WACC* levels used by remuneration committees when determining *EVA* (and include the predetermined *WACC* levels in Stark and Thomas, 1998).[13] [14]

We proxy the amount of *Capital Employed* by the book value of the total assets employed by the company based on accounting data (WorldScope database). More precisely, average annual levels of the book value of total assets employed in year *t* are calculated by taking the average of the year end book value and the book value at the beginning of the year.

The calculated value of accounting based *EVA* follows from the calculated values of *NOPAT*, *WACC* and *Capital Employed* by using equation (3.1). We define the performance measure actually studied, $\Delta EVA = EVA_t - EVA_{t-1}$.

3.2 Relative Total Shareholder Return (*RTSR*)

The most widely accepted goal of for-profit organizations is maximization of value creation (cf. Holström, 1979; Kaplan & Norton, 1996; Baker 1992). In general, value creation is measured in terms of shareholder value creation since shareholders are the residual claimants of the company. Shareholder value creation can be quantified either in dollar terms by means of Market Value Added (*MVA*), or in percentages by means of Total Shareholder Return (*TSR*). Both can be measured relative to a peer group of companies, or in absolute terms. We have selected Relative Total Shareholder Return (*RTSR*) as the relevant measure for the firms' objective function. There are three reasons for doing so.

First, contrary to *MVA*, *(R)TSR* includes dividend payments to shareholders and is therefore independent of whether, and to what extent, firms decide to pay out dividends. Second, absolute measures of value creation such as *MVA* are sensitive to differences in company size, whereas *(R)TSR* is not. Third, due to the inclusion of firms in the treatment group that have introduced *ΔEVA* in various years with possibly different economic market conditions, we use *RTSR* instead of *TSR* to control for these variations in economic conditions. The *RTSR* of firm *j* in year *t* ($RTSR_{j,t}$) is obtained by normalizing the Total Shareholder Return of this firm in year *t* with the *TSR* of the S&P500 index in that year:

---

[12] See e.g. the DEF14A statements of Lyondell, ADC Telecommunications and Whirlpool.

[13] We use a variety of flat *WACC* levels because the *WACC* level may influence the correlation between EVA and the firm's objective in the case of varying levels of capital invested over time due to (de-)investments.

[14] The results shown in what follows have been obtained with a flat *WACC* level of 10%. The results are almost identical when using any other flat WACC level in the range 8-16 percent or when applying equations 3.4 and 3.5 (where r_f is the 5 year federal rate, $r_d = r_f + 2$, and the risk spread is 6.2%). These results are available upon request from the authors.

$$RTSR_{j,t} = \frac{I_{j,t} / I_{j,t-1}}{I_{m,t} / I_{m,t-1}} - 1 \quad (3.6)$$

Here I denotes the stock index provided by the Thompson Datastream database and is calculated as the return on a company's stock, including price changes, dividends payouts, the effects of stock splits, stock shares et cetera.[15] Subscript m refers to the market, j to a firm and t is a year index. The latter refers to the fiscal year that includes the month in which the performance measure has been implemented in the (matched) treatment firm has been adopted as the year t_0.[16]

4 The treatment and control samples

4.1 Firms in the treatment sample

The sample of treated firms consists of 67 NYSE listed companies that have adopted *ΔEVA* to measure and reward (!) the performance of the company's board, see the first column of Table A1. Three sources have been used to identify the population of publicly listed US companies that have adopted *ΔEVA*: (i) a list composed by Stern & Stewart as provided in their company report "The comparative stock market performance of Stern Stewart Clients", (ii) the sample of treatment firms in Wallace (1997),[17] and (iii) a search through the EDGAR database of the US Securities and Exchange Commission (SEC) using keywords "EVA" and "Economic Value Added".[18] The combined lists provide 149 companies that have adopted *ΔEVA* for performance measurement. An analysis of the individual proxy statements of these firms revealed that 74 of them have explicitly adopted *ΔEVA* as a basis for their board members' remuneration. For seven of these companies crucial accounting or market data were lacking, such that a sample of 67 treatment firms resulted. The largest share of treatment firms is active in manufacturing (around 40%; see the *SIC* codes with the first digit equal to 2 or 3 in Table A1).[19]

[15] Stock indices are evaluated over periods of 1, 3, 5 and 7 months such that the influence of the day-by-day volatility of the stock market is smoothened. The results reported in Section 5 pertain to averages over three months, as this is in accordance with general practice (cf. Towers Perrin). Moreover, results appear to be robust to the choice of the number of months.

[16] Because accounting data are annual, they are based on fiscal years that differ between companies. More than half of the companies in our study end their fiscal years in December (57%). The second favorite month is June (18%), the rest is more or less equally distributed.

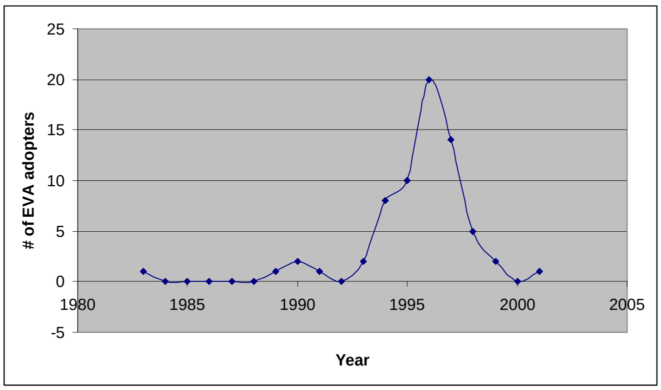
[17] Wallace (1997) focuses on residual income-based compensation plans. A company by company proxy statement assessment revealed whether or not the residual income measure implemented was EVA.

[18] EDGAR, the Electronic Data Gathering, Analysis, and Retrieval system, performs automated collection, validation, indexing, acceptance, and forwarding of submissions by companies and others who are required by law to file forms with the U.S. Securities and Exchange Commission (SEC).

[19] All major *SIC* division structures are represented, except for 'Construction' and 'Public Administration'.

The year of first adopting *ΔEVA* into the board's incentive plan is also retrieved from the proxy statements, see Table A1 in the Appendix and Figure 2 below. In most cases, the proxy statements explicitly mention the date of the decision and the date of implementation. For example: "In July 1996, the Committee approved modifications to the Executive Incentive Plan effective for the fiscal year beginning October 1, 1996." (taken from Becton Dickinson & Co, DEF14A statement Fiscal Year 2007). Figure 2 shows that almost all companies that have adopted *ΔEVA* did so in the nineties, especially in the period 1994-1999.[20] The mean year of implementation is 1995, the median year is 1996.

Figure 2: Distribution of the first adoption of EVA over time



In addition, the proxy statements revealed that 73 percent of the companies in the treatment group use *ΔEVA* as the sole measure in the (short term) incentive plan, as is strongly recommended by Stern Stewart (Stewart, 1991).[21] [22] [23]

4.2 Firms in the control sample

Two control groups are used in our analysis, the second being a subset of the first. The composition of the second (smaller) control group is explained in the next subsection. The first and larger control sample is composed as follows. To each treatment firm we match a control company that has not used

---

[20] One company, Coca Cola, adopted *ΔEVA* already in the early 1980s; two sampled companies adopted *ΔEVA* in 1989 and one in the year 2001, see Table A1 in the Appendix.

[21] Proxy statements mention the use of *ΔEVA* in short term and long term incentive plans. 94 [91] procent of the *ΔEVA* adopting companies use the measure [only] in their short term incentive plan (STI), whereas nine [six] percent use EVA (only) in their long term incentive plan (LTI), in which case *ΔEVA* then determines, for instance, the extent of participation in the company's stock option plan.

[22] The results shown below do not control for the weight attached to *ΔEVA*. The analyses shown have also been performed on the subset of firms that use *ΔEVA* as the sole measure in the board's incentive plan. The results are not any different and available from the authors upon request.

[23] For the 27% of the firms that did not use *ΔEVA* as their sole measure, "implementation of the measure *ΔEVA*" is defined as an increase of the weighting of *ΔEVA* to at least 10% (as stated in the remuneration committee's report in the proxy statement), where *ΔEVA* should not have been used before.

an *RI*-based performance measure in both the period of ten years before (T=I) and after the treatment company has adopted *ΔEVA* (T=II). Matching takes place based on four digit *SIC* codes and company size. As in Wallace (1997) (see also Kleiman, 1999), firms are first matched on the four-digit *SIC* code. Next, the firm that is closest in size, as measured by sales volume in the year prior to the treatment firm's adoption of *ΔEVA* (i.e. year t_0-1), is selected as control. In some cases this matching procedure fails to allocate a suitable control firm, for instance due to an insufficient number of years of data available, large temporal disparities in the fiscal year ending, or large size differences.[24] In these cases, matching takes place on three-digit *SIC* codes or, at a minimum, on two-digit *SIC* codes. The full list of matched control companies is shown in Table A1 in the Appendix.[25]

4.3 Descriptive statistics before treatment

Table 1 shows the descriptive statistics of the treatment and (the larger) control sample in terms of the mean values of *ΔEVA* and *RTSR* before the year t_0 in which the treatment companies have adopted *ΔEVA* as a performance measure. The differences between the treatment and control firms in terms of *RTSR* are insignificant. The values of *ΔEVA* are (significantly) lower in the treatment group than in the control group. The table further shows that the usable sample of matched treatment and control pairs is far below 67 due to the unavailability of pre-treatment time series (mostly in the control sample).

Table 1: Mean values before treatment of *RTSR* and *ΔEVA* in the treatment and control samples

# of years	*RTSR*				*ΔEVA*			
	Treatm Mean	Control Mean	Paired t-test Diff	n	Treatm Mean	Control Mean	Paired t-test Diff	N
k=1	-.037	-.089	n.s.	46	1.76	10.00	n.s.	56
k=2	-.049	-.063	n.s.	52	1.40	5.73	n.s.	56
k=3	-.069	-.062	n.s.	32	0.39	7.12	n.s.	42
k=4	-.041	-.053	n.s.	28	-2.63	4.13	n.s.	38
k=5	-.039	-.031	n.s.	32	-2.44	2.79	*	38
k=6	-.039	-.036	n.s.	34	-1.75	4.35	*	38
k=7	-.042	-.032	n.s.	37	-1.67	4.11	*	38
k=8	-.044	-.031	n.s.	37	-3.17	2.88	**	38
k=9	-.041	-.030	n.s.	37	-3.02	2.96	**	38
k=10	-.041	-.033	n.s.	37	-2.26	1.87	**	38

The first column denotes the number of years *k* included in the calculation of the average values of *RTSR* and *ΔEVA* in period T=I (pre-treatment) and period T=II (post-treatment). The minimum number of available firm-year observations per firm is 1 in the first row, 2 in the second and third row and 3 in all other rows. Each row shows the significance of the difference of the average value of *RTSR* and *ΔEVA* in the treatment and control sample based on a double-sided paired t-test.

[24] The longitudinal nature of the study disqualifies firms that enter and exit the stock market, go bankrupt, are taken over or split up in the period of the study.

[25] Two firms have been used twice as a control firm. Since this number is relatively low we treat these companies as independent observations in the analyses.

"n.s." means that the difference is not significant. A significant difference at the 10% (5%) [1%] level is denoted by * (**) [***].

In order to further evaluate to what extent the treatment and control group differ before treatment with respect to the quantities discussed in Section 2, Table 2 below shows the average values across firms of the before treatment time-series correlation between *ΔEVA* and *RTSR* (Panel A), the average value of the time-series variance of *ΔEVA* (Panel B) and the covariance of *ΔEVA* and *RTSR* before treatment (Panel C). It further shows the results from three one-sided t-tests on a pairwise basis, making use of our matching into treatment-control pairs. The three tests evaluate whether the correlation between *ΔEVA* and *RTSR* before treatment is higher in the treatment sample than in the control sample (Panel A), the variance of *ΔEVA* is lower in the treatment sample than in the control sample (Panel B), and the covariance of *ΔEVA* and *RTSR* before treatment is higher in the treatment sample than in the control sample (Panel C).

Table 2 paints a clear picture in line with the prediction that firms select performance measures based on the pre-treatment level of correlation between the performance measure and the firm's objective. Moreover, the table shows that the before treatment time series variance of the performance measure is indeed lower, on average, for the firms in the treatment group than for the firms in the control group. These differences are insignificant though. The average level of the covariance between *ΔEVA* and *RTSR* is higher in the treatment than in the control group. This is in accordance with the higher mean correlation levels and the (insignificantly) lower variance of *ΔEVA* in the treatment sample.

Table 2: Mean values before treatment of *Corr*(*ΔEVA,RTSR,*), *Var*(*ΔEVA*), and *Cov*(*ΔEVA,RTSR*) in the matched treatment and control samples

# of years	**Panel A: *Corr*(*ΔEVA,RTSR*) before treatment (T=I)**			
	Treatment	Control	Paired t-test: Corr_tr > Corr_contr	
	Mean	Mean	Diff	N
K=5	0.29	-0.02	**	33
K=6	0.31	-0.08	***	33
K=7	0.32	-0.04	***	33
K=8	0.33	-0.04	***	33
K=9	0.31	-0.02	***	33
K=10	0.29	-0.03	***	33
# of years	**Panel B: *Var*(*ΔEVA*) before treatment (T=I)**			
	Treatment	Control	Paired t-test: Var_tr < Var_contr	
	Mean	Mean	Diff	N
K=5	1704.83	2195.55	n.s.	33
K=6	1706.89	2213.65	n.s.	33
K=7	1524.92	2187.35	n.s.	33
K=8	1557.48	2180.64	n.s.	33
K=9	1520.55	2128.02	n.s.	33
K=10	1505.32	2096.45	n.s.	33
# of years	**Panel C: *Cov*(*ΔEVA,RTSR*) before treatment (T=I)**			
	Treatment	Control	Paired t-test: Cov_tr > Cov_contr	
	Mean	Mean	Diff	N
K=5	2.13	0.72	*	33
K=6	2.03	0.22	**	33
K=7	1.91	0.53	**	33
K=8	1.77	0.43	**	33
K=9	1.50	0.64	*	33
K=10	1.50	0.57	*	33

The first column denotes the number of years that has been included in the calculations of the average values in the table. Because the correlations, variances and covariances are based on time-series per company, we use series of at least five years in which a minimum of three non-missing values is observed. Each row shows the significance of the difference of the average values of the correlation, variance and covariance in the respective panels in the treatment and control sample based on a one-sided paired t-test. The alternative hypothesis is that treatment firms have higher values of the correlation and the covariance and lower values of the variance, see Section 2. "n.s." means that the difference is insignificant. A significant difference at the 10% (5%) [1%] level is denoted by * (**) [***].

Table 3 shows the results from a probit analysis that studies the drivers of selection into the treatment sample in a multivariate context. The dependent variable is a dummy variable equal to one for treatment firms and equal to zero for control firms. As potential determinants are included the pre-treatment values of *RTSR*, *ΔEVA*, the time-series correlation between *ΔEVA* and *RTSR,* the time-series variance of *ΔEVA* and the time-series variance of *RTSR*. The latter is included such that all the ingredients of the covariance of *ΔEVA* and *RTSR* are considered as potential determinants. Please note that the detection of distortion does not require that the treatment and control sample are similar with respect to all other possible dimensions.[26] The potential determinants considered here are exactly the ones that may bias the results from the empirical CM tests, see Section 2.

Table 3: Determinants of selection into the treatment sample

Dependent variable = 1 for treatment firms and 0 for control firms		
Variable	*Coefficient*	*Standard error*
ΔEVA_{t0-1}	-0.00315	0.00375
$RTSR_{t0-1}$	0.90096	0.75310
Var(*ΔEVA*)	-0.00004	0.00004
Var(*RTSR*)	-6.05837	5.13541
Corr(*ΔEVA,RTSR*)	0.75294**	0.32900
Constant	0.38116	0.26973
Obs.	83	
Log pseudolikelihood	-52.19	
Wald χ^2 (5)	8.74	
Prob > χ^2	0.12	
Pseudo R^2	0.088	

The t_0-1 values of the independent variables *RTSR* and *ΔEVA* are included, whereas Var(*RTSR*), Var(*ΔEVA*) and Corr(*ΔEVA,RTSR*) are based on the ten years time series t_0-1 to t_0-10. We have performed the same analysis with shorter time series and upon the inclusion of more lags of *RTSR* and *ΔEVA* and found qualitatively the same results. The statistical significance of the coefficients is denoted by stars: the 10% (5%) [1%] levels are indicated * (**) [***] respectively. Robust standard errors are calculated.

The only observed significant determinant of selection into the treatment sample (given a control sample that is similar in terms of industry activity and company size) is the correlation between *ΔEVA* and *RTSR*, whereas var(*ΔEVA*) or the level of *ΔEVA* play no significant role. The higher is the pre-treatment correlation between the performance measure and the company's (presumed) objective function, the more likely the performance measure is used for measurement and reward purposes. This observation indirectly supports the claims by Baker (2002) and Courty and Marschke (2007) that distortion has so far wrongly been measured by the correlation between the value of the performance measure and the 'true' measure of the firm's objective. Table 3 shows that this has not only been the case in empirical research, but also in corporate practice.

The above analysis reveals that self-selection into treatment is strongly related to the measures of association that forms the core of the CM tests of distortion (cf. Section 2); The pre-treatment values of the relevant statistical associations are significantly higher in the treatment group than in the general population of firms. As discussed in Section 2 one could therefore expect regression to the mean, i.e., a decrease in the relevant associations in period T=II vis-à-vis period T=I.

[26] Lovata and Costigan (2002) study some of the characteristics of firms that adopted EVA as a performance measure in their compensation plans. They find that firms with a lower R&D/sales ratio, less insider ownership, and a higher percentage of shares owned by institutional investors are more likely to adopt EVA. Unlike we do, they do not incorporate the observed characteristics of *ΔEVA* in their empirical analysis.

Put differently, decreasing correlation and regression coefficients may not necessarily indicate distortion, but may alternatively point at regression to the mean.

We therefore compose a second control group (taken from the list of firms in the right hand part of Table A1 in the Appendix), that is matched pairwise to the treatment group on the sole basis of the pre-treatment value of the correlation between *ΔEVA* and *RTSR*..[27] The algorithm used is that each pair should consist of a treatment and a control firm whose pre-treatment levels of correlation between *ΔEVA* and *RTSR*. do not differ by more than 0.1 in absolute terms and (for low values) by not more than a factor of 1.5. The resulting sample size reduces to 32 pairs because, given the different distributions of the correlation between *ΔEVA* and *RTSR* in the two initial samples, not all firms can be matched. In particular, a number of observations with very high correlations in the treatment group cannot be included. The same holds for some observations with a very low correlation in the control group.[28]

Matching takes place using the pre-treatment levels of the correlations measured over the period t_0-10 to t_0-1. The correlations based on this time span are very similar to the ones based on the periods starting in t_0-9 to t_0-7. The correlation between these correlations ranges from 0.98 to 0.94. However, time series correlations based on shorter time spans have a lower correlation with those based on the full time span of t_0-10 to t_0-1 (0.89 and 0.83 for the periods starting in t_0-6 and t_0-5, respectively). Therefore, the subsequent analyses are only based on pre- and post-measurement periods ranging from seven to ten years. Tables A2 and A3 in the Appendix show the pre-measurement values for the new sample of matched pairs. (These tables are comparable to Tables 1 and 2.) Differences between the treatment and the second control group appear to be negligible.

In the next section we present the results of the empirical CM tests based on a comparison of the second set of treatment and control firm pairs. Tables A4 to A8 in the Appendix show the comparable results for the pairwise comparison between the treatment and control group where pairing is based on company sizes and SIC codes (i.e. the first control group).

5 Results

In this section we first test for the presence of distortions in *ΔEVA*, using the main components (D.1) through (D.3) of these tests (see Section 2). We then assess whether there is other evidence for agents responding to incentives, based on the remaining components (I.1) and (I.2).

---

[27] Exact propensity score matching is unnecessary here because matching is based on one explanatory variable only.

[28] The usable sample size hardly decreases as compared to the sample sizes reported in Table 2. The observations that are deleted happen to be the ones that pertain to incomplete data series.

5.1 Evidence of distortions and gaming

A first test of distortion is whether the slope coefficient from regressing *RTSR* on (the accounting based measure of) *ΔEVA* decreases upon activation of *ΔEVA* as performance measure in the board's compensation plan. For that purpose we run regressions with *RTSR* as dependent variable and include as regressors (besides a constant): *ΔEVA*, an activation dummy *ACT*, which equals 1 if the company-year observation pertains to a year in which *ΔEVA* has been activated for rewarding the board in the (matched) treatment firm and 0 otherwise, and the multiplication *ΔEVA_ACT* of these two. The coefficient pertaining to this last interaction term is the one of interest, with a significant negative coefficient indicating the presence of distortions in *ΔEVA*. The first column in Table 4 shows the estimated values of this coefficient for the treatment group and the second column the one for the control group. Several different estimates are reported, based on both fixed effects (FE) and random effects (RE-GLS) and on different lengths of the pre-treatment and post treatment periods ranging from seven to ten years (see the different rows). The two columns paint a consistent picture. In the treatment group the interaction term is always significantly negative whereas in the control group it is always insignificantly positive.

To verify whether the coefficient of the interaction term significantly differs between the treatment and the control group we estimate pooled regressions using the observations from the treatment and control group together. As additional regressors we include a dummy variable *TREAT*, equal to one for firms in the treatment group and 0 for those in the control group, and interaction terms of *TREAT* and all other regressors. Here, the coefficient of interest is the one pertaining to the three way interaction term *ΔEVA_ACT_TREAT* and is reported in the third column of Table 4. Again, the result is consistent with *ΔEVA* being a distortionary performance measure: all estimates are significantly negative. In other words, the running down of *ΔEVA* in the treatment group after adopting this measure into the board's compensation plan follows a steeper path than in the control group over the same period.

Although the pooled regressions compare the treatment firms with the control firms, they do not do so on a pairwise basis. The fourth column therefore reports the aggregate results from a more careful difference-in-differences comparison based on matched pairs. First, the same pooled regressions as in the third column are estimated, but now for each pair separately. This results in a series of estimates γ_i for the three-way interaction term *ΔEVA_ACT_TREAT*, with subscript *i* indexing matched pairs. In a second step these estimates are weighted by w_i such that one average estimate $\gamma = \Sigma_i w_i\gamma_i$ results. Following e.g. Cramer (1973) and Leuven and Oosterbeek (2006), the weights w_i are chosen to minimize the variance of the overall estimate γ.[29] The resulting estimate γ is reported in the

[29] In particular, the weights are equal to: $w_i = \dfrac{\mathrm{var}(\gamma_i)^{-1}}{\sum_i \mathrm{var}(\gamma_i)^{-1}}$.

fourth column of Table 4, together with the results from a (one-sided) t-test testing whether $\gamma < 0$. This more careful comparison leads to the same conclusion as before. The slope coefficient of regressing *RTSR* on *ΔEVA* decreases to a larger extent for the treatment firms than for the control firms. This provides evidence for the degradation of *ΔEVA* and thus for the distortion of this performance measure. Table A4 in the Appendix shows comparable results when the other control group is used, where matching is based on company size and SIC codes.

Table 4: Testing for the degradation of the regression slope coefficient (test (D.1))

Dependent variable = *k* years *RSTR* before and after activation								
	Treatment group		Control group		Pooled regression		Diff-in-diff	
	Coeff *Δb*	|t-val|	Coeff. *Δb*	|t-val|	Coeff $\Delta^2 b$	|t-val|	Coeff $\Delta^2 b$	|t-val|
FE, *k=7*	-.001396**	2.05	.000222	0.45	-.001682*	1.91	-.0000623**	2.41
RE, *k=7*	-.001290**	2.03	.000330	0.70	-.001578**	1.98	-.000046**	1.73
FE, *k=8*	-.001263**	1.97	.000340	0.72	-.001603**	2.00	-.0000664**	2.21
RE, *k=8*	-.001190**	1.96	.000347	0.77	-.001495**	1.96	-.0000545**	2.00
FE, *k=9*	-.001351**	2.25	.000257	0.57	-.001606**	2.13	-.0000642**	2.22
RE, *k=9*	-.001155**	2.03	.000204	0.47	-.001333*	1.85	-.0000533**	2.03
FE, *k=10*	-.001394**	2.45	.000158	0.37	-.002961**	2.16	-.0000514**	1.84
RE, *k=10*	-.001195**	2.19	.000130	0.31	-.001309*	1.90	-.0000436*	1.63
# firms	32		32		64		64	

Δb refers to the two-way interaction *ΔEVA_ACT* in the regressions for the treatment and control group separately, $\Delta^2 b$ refers to the three-way interaction *ΔEVA_ACT_TREAT* in the pooled regressions. Statistical significance is denoted by stars: the 10% (5%) [1%] levels are indicated * (**) [***] respectively. The number of stars is sometimes larger than expected based on the t-value due to small sample corrections.

The second test of distortion looks at how the correlation between *ΔEVA* and *RTSR* changes upon the introduction of *ΔEVA* in the board's remuneration function (see (D.2) in Section 2). Table 5 reports the correlation levels between *ΔEVA* and *RTSR* for the treatment and control group separately, both before and after activation of the performance measure. This table also reports the difference between the correlations after and before treatment (i.e. *ΔCorr*) and the difference-in-differences between the treatment and the control group (i.e. $\Delta^2 Corr$). Table A5 in the Appendix shows the results for the treatment and control pairs, where matching is based on firm size and SIC codes.

Table 5: Testing for the degradation of the correlation coefficient (test (D.2))

Corr(ΔEVA,RTSR)	*Treatment*			*Control*			*DID*
	1	2	3 (=2-1)	4	5	6 (=5-4)	7 (=3-6)
	T=I	T=II	*ΔCorr*	T=I	T=II	*ΔCorr*	$\Delta^2 Corr$
k=7	.302	-.039	-.341***	.240	.150	-.090	-.251**
k=8	.270	.013	-.257**	.230	.123	-.107	-.149*
k=9	.251	-.032	-.284***	.252	.150	-.102	-.181**
k=10	.229	-.019	-.248**	.245	.113	-.132*	-.117
# firms			32			32	64

Statistical significance of differences is denoted by stars: the 10% (5%) [1%] levels are indicated * (**) [***] respectively. Significance levels are based on a one-sided paired t-test conform the expectation that the correlations are decreasing and more so for the treatment firms than the control firms, see Section 2.

The results shown in Table 5 (and A5) are supportive of the distortion of the performance measure *ΔEVA* and thus consistent with Table 4. The correlation between *ΔEVA* and *RTSR* decreases significantly in the treatment group, but not in the matched control group. The differences in differences as reported in the final column are large and (marginally) significant. A comparison of Tables 5 and A5 reveals that part of the decrease in the correlation in Table A5 is evidence of regression to the mean, as the differences are slightly smaller in Table 5. However, Table 5 as such clearly supports the distortion of *ΔEVA*.

As discussed in Section 2 distortion can be of two different kinds. First, the performance measure may induce the agent to supply (more) productive effort but in the wrong quantities, because effort is imperfectly valued at the margin. Second, the agent may be motivated to undertake actions that increase the value of the performance measure, but are detrimental to firm value (gaming). To identify whether the distortions are partly of the second kind, the changes in the covariance between *ΔEVA* and *RTSR* upon activation of the performance measure for reward purposes (cf. test (D.3) in Section 2) are calculated. Table 6 (and A6) shows the results.

Table 6: Testing for the degradation of the covariance (test (D.3))

Cov(ΔEVA,RTSR)	*Treatment*			*Control*			*DID*
	1	2	3 (=2-1)	4	5	6 (=5-4)	7 (=3-6)
	T=I	T=II	*ΔCov*	T=I	T=II	*ΔCov*	$\Delta^2 Cov$
k=7	1.337	-.866	-2.203*	1.017	2.222	1.205	-3.408**
k=8	1.254	-.142	-1.396	.910	2.828	1.919#	-3.315**
k=9	.949	-.545	-1.495*	1.224	2.916	1.692#	-3.186**
k=10	.897	-.574	-1.471*	1.111	2.624	1.514#	-2.985**
# firms			32			32	64

Statistical significance of differences is denoted by stars: the 10% (5%) [1%] levels are indicated * (**) [***] respectively. Significance levels are based on a one-sided paired t-test conform the expectation that the covariances are decreasing and more so for the treatment firms than the control firms, see Section 2. Significant differences opposite to the expected direction are denoted by # (# #) [# # #].

The results in Table 6 display a clear pattern in support of gaming: the covariances in the treatment group decrease marginally significantly (third column) and these decreases are significantly larger than those in the control group (final column). Table A6 in the Appendix shows qualitatively similar but less significant results. One caveat applies here, which renders the result in support of gaming somewhat suggestive. Using a decrease in the covariance to detect gaming assumes that overall incentives are not weakened over time, see the formal description of CM test (D.3) in Section 2. Although there is some evidence that this has been the case in general for the time period we are studying (cf. Hall and Liebman, 1998), we are unable to test this assumption for the specific sample of firms considered.

All in all, we conclude that the performance measure *ΔEVA* is distortionary. We find some suggestive evidence that part of this distortion is due to gaming by board members.

5.2 Evidence of responses to incentives

Another intuitive prediction Courty and Marschke derive is that the activation of a performance measure in an agent's compensation plan leads to increased average levels and variances of the performance measure; see predictions (I.1) and (I.2) in Section 2. The driving force is that agents respond to the newly provided incentives (and as with (D.3) the assumption is made that incentives are not weakened over time). Tables 7 and 8 show the results. (Tables A7 and A8 in the Appendix display the results when the other control group is used.)

Table 7: Testing for the increase in average performance (test (I.1))

E(ΔEVA)	*Treatment*			*Control*			*DID*
	1	2	3 (=2-1)	4	5	6 (=5-4)	7 (=3-6)
	T=I	T=II	*ΔE*	T=I	T=II	*ΔE*	$\Delta^2 E$
k=7	1.35	-1.75	-3.10	5.36	-.59	-5.94	2.84
k=8	1.37	-.35	-1.72	3.37	.03	-3.34	1.61
k=9	.96	.40	-.56	3.28	.46	-2.83	2.27
k=10	1.16	1.45	-.28	2.00	.21	-1.79	2.07
# firms			32			32	64

Statistical significance of differences is denoted by stars: the 10% (5%) [1%] levels are indicated * (**) [***] respectively. Significance levels are based on a one-sided paired t-test conform the expectation that the expected (i.e. mean) values are increasing and more so for the treatment firms than the control firms, see Section 2.

Table 8: Testing for the increase in the variance of performance (test (I.2))

Var(ΔEVA)	*Treatment*			*Control*			*DID*
	1	2	3 (=2-1)	4	5	6 (=5-4)	7 (=3-6)
	T=I	T=II	*ΔVar*	T=I	T=II	*ΔVar*	$\Delta^2 Var$
k=7	1202	1974	772*	2356	3544	1188	-416
k=8	1142	1907	765*	2349	3604	1256	-490
k=9	1087	1959	872*	2313	3475	1162	-290
k=10	1075	1922	847*	2279	3597	1318	-470
# firms			32			32	64

Statistical significance of differences is denoted by stars: the 10% (5%) [1%] levels are indicated * (**) [***] respectively. Significance levels are based on a one-sided paired t-test conform the expectation that the variances are increasing and more so for the treatment firms than the control firms, see Section 2.

The third column in Table 7 reveals that (the average of the) mean value of *ΔEVA* over the various numbers of years after adoption of *ΔEVA* is not significantly higher than before. Contrary to expectations, the sign is even negative in all rows. However, the result pertaining to the control group shows that economic circumstances have probably worsened. The final column shows that the decrease in the mean value of *ΔEVA* is consistently larger for the control group than for the treatment group. This DID comparison supports prediction (I.1) of the empirical CM test, although the positive differences are not statistically significant, probably due to the small sample size. The results show the importance of using a control group to correct for (changes in) economic circumstances.

Somewhat ambiguous results are obtained for changes in the variance over time. Table 8 shows that the variance of *ΔEVA* increases (marginally) significantly over time in the treatment group, but not so in the control group. This suggests that activation *ΔEVA* in the board's remuneration package indeed leads to a larger variance in measured performance. In absolute terms the increase in the variance is larger in the control group, however, such that a *DID* comparison leads to insignificant results of the wrong sign. Table A8 in the Appendix reveals that when the other control group is used, the *DID* comparisons have the expected positive signs (but remain insignificant).

Overall, we find some additional, but weak evidence that agents indeed respond to incentives. Average performance and its variance are likely to increase upon activation of *ΔEVA*.

6. Discussion and conclusion

Accounting research has traditionally assessed the quality of performance measures based on the association between the measured value of the performance measure and some indicator of the firm's objective function. Residual income based performance measures – and *EVA* in particular – have been subject to such evaluations, leading to mixed results (e.g. Bacidore et al, 1997; Biddle et al., 1997, 1999; Chen and Dodd, 1997; Feltham et al, 2004; Ismail, 2006; Kyriaris and Anastassis, 2007; O'Byrne, 1996, 1997; Stark and Thomas, 1998; Wallace 1997). More recently, however, it has been shown theoretically that this association is not a relevant criterion to consider for measuring the quality of a performance measure. In particular, Baker (2002) argues that it is not the association between the *levels* of the performance measure and the company's objective function that matters, but rather the association between the *marginal effects* of effort on these two metrics. Unfortunately, the association between these *margins* as a (inverse) measure of the distortiveness of a performance measure is much harder to assess in practice.

In an important recent contribution Courty and Marschke (2007; CM) have been able to develop a suitable empirical test to detect the presence of distortion in performance measures. Basically this test verifies whether the performance measure 'degrades' upon introducing it into the performance contract. This happens to be the case when the statistical association between the performance measure and the firm's objective decreases. The original CM test does not control for self-selection and changes in economic circumstances over time, which is also not required in their particular empirical application. But in more general applications, for instance when assessing the distortion in performance measures that are selected by companies to evaluate and remunerate their board members, one should take selection and timing into account. From a theoretical perspective the most probable drivers of the decision to use a performance measure are related to the association of this measure with the company's objective and/or the variance of the measure. This follows because these quantities are typically viewed as being indicators of distortion and noise, respectively. As a result, the most likely drivers of the adoption of a performance measure are closely related to the statistical measures that CM propose to detect distortion. If these drivers are indeed observed in practice, taking account of self-selection is required to obtain unbiased tests of distortion in performance measures.

In this paper we apply the CM test to the performance measure *EVA*, a Residual Income based performance measure widely used in corporate practice. We adapt the test to cope with self-selection and timing decisions. In particular, we employ a difference-in-differences approach, where the control group consists of a matched sample of suitably selected firms. By comparing the changes in test outcomes before and after adopting the performance measure (by treatment firms) across the treatment and control groups we account for timing issues.

Overall, our test results indicate that *EVA* is a distortionary performance measure. Both the slope coefficient of regressing *RTSR* on *ΔEVA* and the correlation coefficient between these two quantities decrease to a larger extent in the treatment group than in the control group. Our finding that the covariance between *RTSR* and *ΔEVA* decreases to a significantly larger extent for the treatment firms, supports gaming. This suggests that the distortions detected are partly due to gaming. Our results are less conclusive here, however, because the covariance test requires that overall incentives have not weakened over time, on which we have no hard data.

Given the theoretical results of Rogerson (1997) and Reichelstein (1997, 2000) that, under certain assumptions, rewarding managers on the basis of a residual income based performance measure like *EVA* will give them correct investment incentives,[30] our empirical finding that *EVA* is a distortionary performance measure that can be gamed is (perhaps) somewhat surprising. It immediately raises the question what kind of decisions and/or activities board members can undertake that are consistent with maximizing the value of *EVA*, but are misaligned with shareholder value creation. Existing research is of little help in clearly identifying such distortionary actions, but does provide some tentative indications. Using a matched pair approach, Wallace (1997) compares the behavior of managers in firms that introduce residual income-based performance rewards like *EVA* to firms that use more traditional earnings-based performance rewards. He finds (among other things) that relative to non-adopters, *EVA* adopters dispose of more assets and decrease their new investments. As noted by Wallace (1997, p. 287), it is difficult to establish whether the observed reduction in net investments is value-increasing; it may well be the case that due to the *EVA*-based rewards managers actually under-invest in projects that render a positive Net Present Value (*NPV*). Similar remarks apply to the two other main findings of Wallace, viz. managers who are rewarded on the basis of *EVA* increase their share repurchases and dividend payouts and use their assets more intensely. Misalignment with value creation might be the result if dividend payment and share repurchase levels are increased beyond the optimal amount, or if assets are used too intensively. This could be the underlying cause of the distortionary behavior induced by *EVA*-based rewards that our study reveals.

The other potential causes of the distortiveness of *EVA* relate to 'short termism'. Because *EVA* is a single period performance measure it has the potential drawback that it induces managers to take improper account of the long-term (Bromwich and Walker, 1998). The Stern Stewart EVA

[30] The intuition behind the theoretical optimality of rewarding managers on the basis of residual income is that, by choosing an appropriate depreciation schedule, investment costs can be spread out over the investment's life time exactly proportional to the benefits. Residual income then reflects the value created by the manager at any given point in time, providing her with the same (optimal) incentive to invest in every period. The Rogerson (1997) and Reichelstein (1997) models focus on investment incentives and effectively ignore the moral hazard problem, i.e. managers need not be induced to take appropriate effort. (Bromwich and Walker (1998, p. 412ff.) provide an insightful discussion of the properties of *EVA* like reward systems these two models do not explore.) Reichelstein (2000) extends the earlier models with an effort dimension and shows that the agency costs of inducing both efficient investment and efficient effort are lower for residual income than for performance measures based on realized cash flows only (i.e. that do not incorporate a capital charge).

approach tries to address the various problems that this may lead to in three different ways (cf. O'Hanlon and Peasnell, 1998): (i) by making adjustments to the accounting based elements of Residual Income (cf. equation (3.2)), (ii) by using non-zero *EVA* benchmarks in performance evaluation and (iii) by separating the award and payment of bonuses. Concerning the first, one important purpose of the adjustments is to discourage earnings management, i.e. to limit the opportunities managers have to allocate *EVA* across different periods. But as O'Hanlon and Peasnell (1998, p. 432) argue, it is unclear whether the adjustments made will solve this gaming problem.[31] More generally they note that, even after all adjustments are made, a positive (negative) level of *EVA* not necessarily indicates value creation (destruction). This necessitates the second adaptation of using a non-zero *EVA* benchmark, which in turn leads to the difficult issue of how this benchmark is set. If set inappropriately, dysfunctional behavior may be induced.

The third element of using a 'bonus bank' entails that bonuses are not immediately paid out in cash, but smoothed out over a period of three years. The main reason for doing so is again to avoid potentially dysfunctional short-term behavior of managers – like engaging in *EVA* accelerating activities – which may be induced even after all the recommended adjustments to *EVA* have been made. Although the bonus bank scheme is likely to reduce the worst of such 'short termism' effects, it may not eliminate all dysfunctional effects (cf. Otley, 1999, p. 373). In particular, managers with short-term horizons may have an incentive to avoid negative *EVA* projects even if these projects are profitable in the long run (cf. Brickley et al, 2004).

Although suggestive, in the absence of conclusive empirical data the above discussion of potentially distortionary actions is mere speculation. Therefore, more research is required to identify exactly what ingredient of *EVA* causes the distortiveness of this measure that is widely used as a basis for performance pay of managing board members. Nevertheless and in the mean time, our empirical results indicate that the incentives provided by the performance measure *EVA* are suboptimal.

[31] As Zimmerman (1997, p. 107) notes, the adjustments themselves may also induce gaming. One example he provides concerns the reliance of *EVA* on managers' estimates of bad debt expenses. By underestimating these expenses, managers can overstate their performance and thereby boost their bonuses. Other examples include delaying inventory writedowns or using long lives for estimating depreciation schedules.

References

Anthony, R., 1973. "Accounting for the Cost of Equity", *Harvard Business Review* 51, 88-102.

Anthony, R., 1982a. "Equity Interest - Its Time has Come", *Journal of Accountancy* 154, 76-93.

Anthony, R., 1982b. "Recognizing the Cost of Interest on Equity", Harvard Business Review 60, 91-96.

Bacidore, J., Boquist, J., Milbourn, T., and A. Thakor, 1997, "The Search for the Best Financial Performance Measure", *Financial Analysts Journal* 53, 11-20.

Baker, G.P., 1992. "Incentive Contracts and Performance Measurement", *Journal of Political Economy* 100(3), 598-613.

Baker, G., 2002. "Distortion and Risk in Optimal Incentive Contracts", *The Journal of Human Resources* 37(4), 728-751.

Biddle, G., G. Seow and A. Siegel, 1995. "Relative versus Incremental Information Content" *Contemporary Accounting Research* 12, 1-23.

Biddle, G., R. Bowen and J. Wallace, 1997. "Does EVA beat Earnings? Evidence on Associations with Stock Returns and Firm Values", *Journal of Accounting and Economics* 24, 301-336.

Biddle, G., R. Bowen and J. Wallace, 1999. "Evidence on EVA", *Journal of Applied Corporate Finance* 12, 69-79.

Bowen, R., M. Johnson, T. Shevlin and D. Shores, 1989. "Informational Efficiency and the Information Content of Earnings during the Stock Market Crash of 1987", *Journal of Accounting and Economics* 11, 225-254.

Brickley, J.A., Smith, C.W. and J.L. Zimmerman, 2004, *Managerial Economics and Organizational Architecture*, 3rd edition, McGraw-Hill, New York.

Bromwich, M. and M. Walker, 1998, "Residual Income Past and Future", *Management Accounting Research* 9, 391-419.

Chen, S. and J.L. Dodd, 1997, "Economic Value Added (EVA®): An Empirical Examination of a new Corporate Performance Measure", *Journal of Managerial Issues* 9, 318-333.

Coles, J., N. Daniel and L. Laeven, 2006. "Managerial Incentives and Risk Taking", Journal of Financial Economics 79(2), 431-468.

Courty, P. and G. Marschke, 2007. "A General Test for Distortions in Performance Measures", forthcoming in *The Review of Economics and Statistics.*

Cramer, J.S., 1973. "Interaction of Income and Price in Consumer Demand", *International Economic Review* 14(2), 351-363.

Datar, S., S. Kulp and R. Lambert, 2001. "Balancing Performance Measures", *Journal of Accounting Research* 39(1), 75-92.

Fama, E. and K French, 1996. "CAPM is Wanted, Dead or Alive", *Journal of Finance* 51, 1947-1958.

Feltham, G. and J. Xie, 1994. “Performance Measure Congruity and Diversity in Multi-Task Principal Agent Relations”, *The Accounting Review* 69, 429-453.

Feltham, G, G. Issac, C. Mbagwu and G. Vaidyanathan, 2004. “Perhaps EVA Does Beat Earnings — Revisting Previous Evidence”, *Journal of Applied Corporate Finance* 16, 83-88.

Gebhardt, W., C. Lee and B. Swaminathan, 2001. “Toward an Implied Cost of Capital”, *Journal of Accounting Research* 39, 135-176.

Hall, B.J. and J.B. Liebman 1998. “Are CEOs Really Paid like Bureaucrats?”, *Quaterly Journal of Economics* 113, 653-691.

Hamilton, R.., 1777. *An Introduction to Merchandize*, Edinburgh.

Holmstrom, B., 1979. “Moral Hazard and Observability”, *The Bell Journal of Economics* 10, 74-91.

Holmstrom, B. and P. Milgrom, 1991. “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design”, *Journal of Law, Economics, and Organization* 7, 24-52.

Ismail, A., 2006, “Is Economic Value Added More Associated with Stock Return than Accounting Earnings? The UK Evidence”, *International Journal of Managerial Finance* 2, 343-353.

Kaplan, R. and D. Norton, 1996. *The Balanced Scorecard* Boston: Harvard Business School Press.

Kleiman, R.T., 1999, “Some New Evidence on EVA Companies”, *Journal of Applied Corporate Finance* 12, 80-91.

Kyriaris, D. and Ch. Anastassis, 2007, “The Validity of the Economic Value Added Approach: An Empirical Application”, *European Financial Management* 13, 71-100.

Leuven, E. and H. Oosterbeek, 2006. “*The* R*esponsiveness of* T*raining* P*articipation to* T*ax* D*eductibility*”, working paper.

Lovata, L.M. and M.L. Costigan, 2002, “Empirical Analysis of Adopters of Economic Value Added”, *Management Accounting Research* 13, 215-228.

Marshall, A., 1890. *Principles of Economics* The Macmillan Press Ltd, London, New York.

O’Byrne, S.F., 1996, “EVA and Market Value”, *Journal of Applied Corporate Finance* 9, 116-126.

O’Byrne, S.F., 1997, “EVA and Shareholder Return”, *Financial Practice and Education* 7, 50-54.

O’Hanlon, J. and K. Peasnell, 1998, “Wall Street’s contribution to Management Accounting: The Stern Stewart EVA® Financial Management System”, *Management Accounting Research* 9, 421-444.

Otley, D., 1999, “Performance Management: A Framework for Management Control Systems Research”, *Management Accounting Research* 10, 363-382.

Prendergast, C., 2002. “The Tenuous Tradeoff between Risk and Incentives”, *Journal of Political Economy* 110, 1071-1102.

Reichelstein, S., 1997, “Investment Decisions and Managerial Performance Evaluation”, *Review of Accounting Studies* 2, 157-180.

Reichelstein, S., 2000, “Providing Managerial Incentives: Cash Flows versus Accrual Accounting”, *Journal of Accounting Research* 38, 243-269.

Rogerson, W.P., 1997, "Intertemporal Cost Allocation and Managerial Investment Incentives: A Theory Explaining the Use of Economic Value Added as a Performance Measure", *Journal of Political Economy* 105, 770-795.

Solomons, D., 1965. Division *Performance: Measurement and Control*. Financial Executives Research Foundation, New York.

Stark, A. and H. Thomas, 1998. "On the Empirical Relationship between Market Value and Residual Income in the UK", *Management Accounting Research* 8(4), 445-460.

Stern Stewart EVA Roundtable, 1994. *Journal of Applied Corporate Finance* 7, 46-70.

Stern, J., G. Stewart and D. Chew, 1995. "The EVA Financial Management System", *Journal of Applied Corporate Finance* 8, 32–46.

Stewart, G., 1991. *The Quest For Value*. Harper Business, New York.

Wallace, J., 1997. "Adopting Residual Income-Based Compensation Plans: Do You Get WhatYou Pay For?" *Journal of Accounting and Economics,* 24, 275-300.

Zimmerman, J.L., 1997, "EVA and Divisional Performance Measurement: Capturing synergies and Other Issues", *Journal of Applied Corporate Finance* 10, 98-109.

Appendix Table A1 Companies in the treatment and control group

Name	Ticker	Time of implementation	SIC code	Paired Company	Ticker
Acxiom Corp.	AIH	November-94	7374	Total System Services Inc	TSS
ADC Telecommunications Inc	ACXM	November-96	3661	Arris Group Inc	ARRS
Jarden Corp.	ADCT	January-94	4783	American Greetings Corp.	AM
Ball Corp.	AHIZ	August-91	3411	Crown Holdings Inc	CCK
Bausch & Lomb Inc	ALC/JAH	August-96	3851	Oakley Inc	OO
Becton Dickinson & Company	ILA	October-95	3841	3M Company	MMM
Best Buy Company Inc	ACKH	March-99	5731	Circuit City Inc	CC
Officemax Incorporated	AVY	January-95	5943	Staples Inc	SPLS
Bowater Inc	BLL	January-96	2621	Glatfelter	GLT
Briggs & Stratton Corp.	BOL	January-90	3510	Caterpillar Inc	CAT
Centura Banks Inc.	BDX	January-94	6022	Doral Financial Corp.	DRL
Coca Cola Company	BBY	July-83	2086	Pepsico Inc	PEP
Columbus McKinnon Corp.	OMX	April-97	3536	Alamo Group Incorporated	ALG
Crane Company	BOW	July-90	3490	Parker-Hannifin Corp.	PH
ELI Lilly & Company	BGG	January-95	2834	Abbott Laboratories Inc	ABT
Federal Mogul Corp.	CBC	May-97	3713	Dana Corp.	DCNAQ
Flemming Companies Inc	KO	February-94	5141	Performance Food Group Company	PFGC
GC Companies INC	CMCO	April-96	7832	Carmike Cinemas Inc	CKEC
Georgia Pacific Corp.	CR	April-94	2400	Glatfelter	GLT
Guidant Corp.	FDMLQ	January-95	3845	Fischer Imaging Corp.	FIMG
Miller (Herman) Inc	FLMI	March-96	2522	Flexsteel Industries Inc	FLXS
International Multifoods Corp.	GCCX	March-01	5149	McCormick & Company Inc	MKC
Penney JC Company Inc	GCO	October-97	5311	Federated Department Stores Inc	FD
Johnson Outdoors Inc	GP	February-96	3792	Fleetwood Enterprises	FLE
Great Plains Energy Inc	GLK	July-95	4911	Ohio Power Company	OHIPN
Manitowoc Company Inc	GXPP	March-93	3531	Alico Inc	ALCO
Material Sciences Corp.	GDT	September-97	3479	Lindsay Corporation	LNN
Millenium Chemicals Inc.	ICPXQ	October-96	2899	PPG Industries Inc	PPG
Touch America Holdings Inc	IMC	February-97	4899	Ohio Power Company	OHIPN
Noble Corp.	ITT	July-97	1381	Abraxas Petroleum Corp.	ABP
Olin Corp.	JOUT	January-96	3341	Rockwood Holdings Inc	ROC
Perkinelmer Inc	KSWS	August-94	3841	Applera Corp.	ABI
Polaroid Corp.	KBAL	August-96	3827	Cooper Companies Inc	COO
Donnelley RR & Sons Company	LLY	November-94	2754	Interpublic Group Companies Inc	IPG
Ryder System Inc	LYO	January-97	7513	Amerco	UHAL
SVB Financial Group	MAG	January-96	6021	Alabama National Bancorporation	ALAB
SPX Corp.	MTW	January-96	3823	Ametek Inc	AME
Standard Motor Products Inc	MRO	January-98	3694	Motorcar Parts Of America	MPAA
Tenet Healthcare Corp.	TUG	January-97	8062	Universal Health Services Inc	UHS
Tupperware Brands Corp.	MSC	June-96	3089	Aptargroup Inc	ATR
Vulcan Materials Corp.	MDII	January-96	1422	MDU Resources Group Inc	MDU
Webster Financial Corp.	MZ	January-97	6021	Downey Financial Corp.	DSL
Whirlpool Corp.	MCH	January-97	3630	Centex Corp.	CTX
Grainger WW Inc	HLHR	July-89	5000	Allied Waste Industries Inc	AW
Milacron Inc	MUR	January-93	3559	Wellco Enterprises Inc	BRKS
Maritrans Partners	NE	January-94	4499	Sea Containers Limited	SCRB
Ablest Inc	OLN	January-95	7361	National Technical Systems Inc	NTSC
Allied Holdings Inc	JCP	January-97	4213	Covenant Transport Inc	CVTI
MDI Inc	PKI	January-97	3829	Mechanical Technology Inc	MKTY
Aquila Inc	PRD	January-95	4922	El Paso Corp.	EP
Armstrong Holdings Inc	RRD	January-95	3996	Interface Inc	IFSIA
Avery Dennison Corp.	R	January-96	2891	Cytec Industries	CYT
Genesco Inc	SVM	February-99	5661	Brown Shoe Inc	BWS
Great Lakes Chemical Corp.	SPW	January-98	2819	Church & Dwight Company Inc	CHD
Inacom Corp.	SMP	January-95	5045	En Pointe Technology Inc	ENPT
ITT Corporation	SIVB	January-98	3812	General Dynamics	GD
K-Swiss Inc	TEK	June-96	5139	Kenneth Cole Productions Inc	KCP
Kimball International Inc	THC	June-96	2517	Ethan Allen Interiors Corp.	ETH
Lyondell Chemical Company	TEN	January-95	2869	Dow Chemicals Company	DOW
Magnetek Inc	MTP	June-96	3679	BTU International	BTUI
Marathon Oil Corp.	TUP	January-96	2911	Hess Corp.	HES
Murphy Oil Corp.	VMC	January-96	2911	Frontier Oil Corp. Commerce	FTO
Servicemaster Company	GWW	January-98	0782	Central Garden & PET Company	CENT
Tektronix Inc	WBS	June-95	3825	COHU Inc	COHU
Tenneco Inc	WHR	January-97	3714	Borgwarner Inc	BWA
Wolohan Lumber Corp.	WHLN	January-97	5211	Building Materials Holdings Corp.	BLG
Worthington Industries Inc	WOR	June-98	3312	AK Steel Holdings	AKS

Table A2: Mean values before treatment of *RTSR* and *ΔEVA* in the treatment and the second matched control sample (based on *Corr(ΔEVA,RTSR)*

# of years	*RTSR*				*ΔEVA*			
	Treatm	Control	Paired t-test		Treatm	Control	Paired t-test	
	Mean	Mean	Diff	N	Mean	Mean	Diff	n
k=7	-0.050	-0.022	n.s.	31	-1.340	5.226	*	30
k=8	-0.052	-0.019	n.s.	31	-1.316	3.165	n.s.	30
k=9	-0.046	-0.021	n.s.	31	-1.753	3.192	n.s.	30
k=10	-0.047	-0.022	n.s.	31	-1.540	1.861	n.s.	30

The first column denotes the number of years *k* that has been included in the calculation of the average values of *RTSR* and *ΔEVA* in period T=I (pre-treatment) and period T=II (post-treatment). The minimum number of available firm-year observations per firm is 3 in all rows. Each row shows the significance of the difference of the average value of *RTSR* and *ΔEVA* in the treatment and control sample based on a double-sided paired t-test. "n.s." means that the difference is not significant. A significant difference at the 10% (5%) [1%] level is denoted by * (**) [***].

Table A3: Mean values before treatment of *Corr*(*ΔEVA,RTSR,*), *Var*(*ΔEVA*), and *Cov*(*ΔEVA,RTSR*) in the second matched treatment and control samples (based on *Corr*(*ΔEVA,RTSR,*))

# of years	**Panel A: *Corr*(*ΔEVA,RTSR*) before treatment (T=I)**			
	Treatment	Control	Paired t-test: Corr_tr > Corr_contr	
	Mean	Mean	Diff	n
K=7	0.30	0.24	n.s.	32
K=8	0.27	0.23	n.s.	32
K=9	0.25	0.25	n.s.	32
K=10	0.23	0.24	n.s.	32
# of years	**Panel B: *Var*(*ΔEVA*) before treatment (T=I)**			
	Treatment	Control	Paired t-test: Var_tr < Var_contr	
	Mean	Mean	Diff	n
K=7	1201.61	2356.09	n.s.	32
K=8	1141.92	2348.74	n.s.	32
K=9	1086.85	2313.08	*	32
K=10	1074.53	2279.40	*	32
# of years	**Panel C: *Cov*(*ΔEVA,RTSR*) before treatment (T=I)**			
	Treatment	Control	Paired t-test: Cov_tr > Cov_contr	
	Mean	Mean	Diff	n
K=7	1.34	1.01	n.s.	32
K=8	1.25	0.91	n.s.	32
K=9	0.95	1.22	n.s.	32
K=10	0.90	1.11	n.s.	32

The first column denotes the number of years that has been included in the calculations of the average values in the table. Each row shows the significance of the difference of the average values of the correlation, variance and covariance in the respective panels in the treatment and control sample based on a one-sided paired t-test. Here the alternative hypothesis is that treatment firms have higher values of the correlation and the covariance, and lower values of the variance, see Section 2. "n.s." means that the difference is insignificant. A significant difference at the 10% (5%) [1%] level is denoted by * (**) [***].

Table A4: Testing for the degradation of the regression slope coefficient (test (D.1)); matching based on company size and SIC codes

Dependent variable = *k* years *RSTR* before and after activation								
	Treatment group		Control group		Pooled regression		Diff-in-diff	
	Coeff Δb	|t-val|	Coeff. Δb	|t-val|	Coeff $\Delta^2 b$	|t-val|	Coeff $\Delta^2 b$	|t-val|
FE, *k=7*	-.000982*	1.92	.000408	0.83	-.001391**	1.96	.0000040	0.12
RE, *k=7*	-.000934*	1.92	.000485	1.02	-.001412**	2.08	.0000289	0.10
FE, *k=8*	-.000827*	1.73	.000500	1.09	-.001328**	2.00	-.0000076	0.25
RE, *k=8*	-.000776*	1.69	.000485	1.08	-.001250*	1.94	-.0000083	0.33
FE, *k=9*	-.000866*	1.94	.000445	1.02	-.001303**	2.09	.0000080	0.26
RE, *k=9*	-.000866*	1.92	.000421	0.98	-.001232**	2.03	.0000029	0.12
FE, *k=10*	-.000818*	1.93	.021693	0.71	-.001109*	1.87	.0000051	0.18
RE, *k=10*	-.000819**	2.00	.000275	0.68	-.001085*	1.88	.0000021	0.08
# firms	66		66		132		132	

Δb refers to the two-way interaction *ΔEVA_ACT* in the regressions for the treatment and control group separately, $\Delta^2 b$ refers to the three-way interaction *ΔEVA_ACT_TREAT* in the pooled regressions. Statistical significance is denoted by stars: the 10% (5%) [1%] levels are indicated * (**) [***] respectively. The number of stars is sometimes larger than expected based on the t-value due to small sample corrections.

Table A5: Testing for the degradation of the correlation coefficient (test (D.2)); matching based on company size and SIC codes

Corr(ΔEVA,RTSR)	*Treatment*			*Control*			*DID*
	1	2	3 (=2-1)	4	5	6 (=5-4)	7 (=3-6)
	T=I	T=II	*ΔCorr*	T=I	T=II	*ΔCorr*	$\Delta^2 Corr$
k=7	.323	.140	-.184**	.028	.039	.011	-.231*
k=8	.310	.152	-.158**	.023	.043	.020	-.229*
k=9	.295	.135	-.161**	.063	.063	-.001	-.205*
k=10	.282	.155	-.127*	.059	.032	-.026	-.190*
# firms			50			41	33

Statistical significance of differences is denoted by stars: the 10% (5%) [1%] levels are indicated * (**) [***] respectively. Significance levels are based on a one-sided paired t-test conform the expectation that the correlations are decreasing and more so for the treatment firms than the control firms, see Section 2. Since the numbers in columns 1-3 and 4-6 are based on larger sample sizes than the matched comparison in column 7, the difference in column 7 cannot be derived from the numbers in columns 1-6.

Table A6: Testing for the degradation of the covariance (test (D.3)); matching based on company size and SIC codes

Cov(ΔEVA,RTSR)	*Treatment*			*Control*			*DID*
	1	2	3 (=2-1)	4	5	6 (=5-4)	7 (=3-6)
	T=I	T=II	*ΔCov*	T=I	T=II	*ΔCov*	$\Delta^2 Cov$
k=7	1.759	1.632	-.126	-.043	1.509	1.552#	-.104
k=8	1.600	1.758	.158	-.126	2.049	2.175# #	-.698
k=9	1.386	1.513	.126	.107	2.135	2.028# #	-.676
k=10	1.332	2.101	.769	.032	1.785	1.753#	-.471
# firms			50			41	33

Statistical significance of differences is denoted by stars: the 10% (5%) [1%] levels are indicated * (**) [***] respectively. Significance levels are based on a one-sided paired t-test conform the expectation that the covariances are decreasing and more so for the treatment firms than the control firms, see Section 2. Significant differences opposite to the expected direction are denoted by # (# #) [# # #]. Since the numbers in columns 1-3 and 4-6 are based on larger sample sizes than the matched comparison in column 7, the difference in column 7 cannot be derived from the numbers in columns 1-6.

Table A7: Testing for the increase in average performance (test (I.1)); matching based on company size and SIC codes

E(ΔEVA)	*Treatment*			*Control*			*DID*
	1	2	3 (=2-1)	4	5	6 (=5-4)	7 (=3-6)
	T=I	T=II	*ΔE*	T=I	T=II	*ΔE*	$\Delta^2 E$
k=7	1.32	-1.31	-2.63	3.85	-.03	-3.88#	2.31
k=8	.53	-.05	-.58	2.78	-.07	-2.85	3.09
k=9	.57	1.19	.62	2.70	.73	-1.97	3.65
k=10	.98	2.34	1.36	1.98	.67	-1.31	2.98
# firms			62			**??**	**??**

Statistical significance of differences is denoted by stars: the 10% (5%) [1%] levels are indicated * (**) [***] respectively. Significance levels are based on a one-sided paired t-test conform the expectation that the expected (i.e. mean) values are increasing and more so for the treatment firms than the control firms, see Section 2. Significant differences opposite to the expected direction are denoted by # (# #) [# # #]. Since the numbers in columns 1-3 and 4-6 are based on larger sample sizes than the matched comparison in column 7, the difference in column 7 cannot be derived from the numbers in columns 1-6.

Table A8: Testing for the increase in the variance of performance (test (I.2)); matching based on company size and SIC codes

Var(ΔEVA)	*Treatment*			*Control*			*DID*
	1	2	3 (=2-1)	4	5	6 (=5-4)	7 (=3-6)
	T=I	T=II	*ΔVar*	T=I	T=II	*ΔVar*	$\Delta^2 Var$
k=7	1279	2595	1315**	2400	2866	467	848
k=8	1252	2364	1113***	2394	2955	561	594
k=9	1205	2585	1380***	2367	2876	509	1071
k=10	1168	2683	1514***	2340	2963	623	1003
# firms			50			42	33

Statistical significance of differences is denoted by stars: the 10% (5%) [1%] levels are indicated * (**) [***] respectively. Significance levels are based on a one-sided paired t-test conform the expectation that the variances are increasing and more so for the treatment firms than the control firms, see Section 2. Since the numbers in columns 1-3 and 4-6 are based on larger sample sizes than the matched comparison in column 7, the difference in column 7 cannot be derived from the numbers in columns 1-6.