Testing for Distortions in Performance Measures
An Application to Residual Income Based Measures like Economic Value Added

Sloof, Randolph; Van Praag, Mirjam

Document Version
Final published version

Publication date:
2014

Citation for published version (APA):
Sloof, R., & Van Praag, M. (2014). *Testing for Distortions in Performance Measures: An Application to Residual Income Based Measures like Economic Value Added*. Tinbergen Institute. Tinbergen Institute Discussion Papers No. 14-056/V http://www.tinbergen.nl/discussionpaper/?paper=2314

TI 2014-056/V
Tinbergen Institute Discussion Paper

Testing for Distortions in Performance Measures: An Application to Residual Income Based Measures like Economic Value Added

Randolph Sloof
*Mirjam van Praag**

Faculty of Economics and Business, University of Amsterdam, and Tinbergen Institute.

** Copenhagen Business School, Denmark.*

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

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

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Testing for Distortions in Performance Measures:

An Application to Residual Income Based Measures like Economic Value Added

Accepted for publication in the *Journal of Economics and Management Strategy*

Randolph Sloof*

University of Amsterdam and Tinbergen Institute

Mirjam van Praag

Copenhagen Business School

University of Amsterdam and Tinbergen Institute

ABSTRACT: Distorted performance measures in compensation contracts elicit suboptimal behavioral responses that may even prove to be dysfunctional (gaming). This paper applies the empirical test developed by Courty and Marschke (2008) to detect whether the widely used class of Residual Income based performance measures —such as Economic Value Added (*EVA*)— is distorted, leading to unintended agent behavior. The paper uses a difference-in-differences approach to account for changes in economic circumstances and the self-selection of firms using *EVA*. Our findings indicate that *EVA* is a distorted performance measure that elicits the gaming response.

JEL-codes: D21, G35, J33, L21, M12, M40, M52

Keywords: Residual Income, Economic Value Added, distortion, performance measurement, incentive compensation

---

* This version: November 2012. Corresponding author: r.sloof@uva.nl; Roetersstraat 11, 1018 WB Amsterdam, the Netherlands. Tel: +31 20 5255241; fax: +31 20 5254254. We gratefully acknowledge the research assistance of Dennis Schoenmakers and comments by George Baker and Luis Garicano. The usual disclaimer applies. We thank the (associate) editor and an anonymous referee for their excellent suggestions.

1. INTRODUCTION

Understanding the incentive implications of performance measures used in remuneration contracts has become vital since the rapid introduction of performance-based compensation schemes in the last three decades. The classic principal-agent model acknowledges that the noise of a given performance measure determines its suitability for use in compensation contracts. The "noisier" the 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, but rather its "distortion". Baker defines distortion inversely as the extent to which the effect of effort on measured performance is aligned with the effect of effort on the firm's objective function.[1]

If a performance measure is "distorted", then agents can increase their performance outcomes in two ways: by engaging in productive activities that are intended by the principal or by engaging in unintended actions that are easier or cheaper from the agent's perspective. Performance pay based on a distorted performance measure will thus motivate agents to put costly effort into 'cheap' activities (increasing measured performance) to the possible detriment of organizational value. Such unproductive efforts are typically referred to as 'gaming efforts' (see Courty and Marschke, 2008).

Performance measures are often assessed and chosen based on their correlation with firm value (see Baker, 2002; Courty and Marschke, 2008). However, some measures that seem to be informative ex ante about the performance of the company become less informative when used for incentive purposes, revealing their distortion. Implementing a compensation scheme based on such distorted measures undermines the association between the measure and the value of the company. Put differently, the mere use of a distorted performance measure in compensation degrades its quality as a measure of intended performance. The degree of (potential) degradation depends on the extent to which the measure is distorted. Thus, the suitability of a performance measure cannot

[1] Theoretical multi-tasking models show an inverse relationship between the distortion of the performance measure and the efficient incentive intensity (see Holmstrom and Milgrom, 1991; Feltham and Xie, 1994; Datar et al., 2001; Baker, 2002).

be measured by the ex ante correlation between the measure and the value of the company, as this correlation disregards the distortion not yet revealed.

An example involves the distortion of the performance measure "client satisfaction" when it is used in compensation schemes for sales employees. Client satisfaction, used as a performance measure, may provide valuable information about a sales employee's contribution to company performance. Using the measure as a basis for performance pay, however, creates an incentive for a sales employee to increase client satisfaction through 'cheap' effort: selling at low prices, for example, or even giving products away for free. The performance measure "client satisfaction" thus becomes degraded: it is less useful than originally expected as a basis for performance pay.

Courty and Marschke (2008) developed an empirical test to detect distortion. Their test, based on a simple theoretical framework, assesses how a performance measure degrades when it is first used in a compensation contract or when its use in such a contract is intensified. The focus is on measuring the *change* in the association between the performance measure and organizational value due to an increase in the weight of the measure in the compensation scheme. A negative change indicates degradation of the quality of the measure, revealing a distortion. Courty and Marschke apply their test to performance measures introduced in the course of a natural experiment among agencies managing the US Governmental Job Training Partnership Act. They find weak support for the distortion of these measures.

This paper applies the Courty and Marschke (2008) approach to test for distortions in the performance measure Residual Income (*RI*), a key indicator of corporate performance. More specifically, we collected a sample of firms that have introduced the Residual Income based performance measure Economic Value Added (*EVA)* in the remuneration contract of executive board members. *EVA* was developed and copyrighted by Stern Stewart & Co. Unlike *RI*, *EVA* cannot be measured solely on the basis of firms' accounting data. It differs from Residual Income due to some —for researchers opaque— discretionary and standard adjustments. In our empirical analysis we measure *RI* for the firms in our sample and refer to this as the *accounting based value* of *EVA*.[2]

Unlike Courty and Marschke, the application in this paper does not concern a natural experiment, but reviews the experience of listed (US-based) firms that introduced this popular performance measure to reward management performance, mostly in the mid-nineties. This non-experimental approach requires adapting the empirical Courty and Marschke test in order to address self-selection and to account for possible (systematic) changes in economic circumstances in the period studied. To this end, we apply a difference-in-differences approach and compare firms that adopted *EVA* with their non-adopter counterparts, before and after the boards of the former firms selected *EVA* as the performance measure in their employee compensation schemes. The results of this study indicate that the accounting based version of the company performance measure *EVA* is a distorted performance measure that can be gamed.

Our findings should not be interpreted as showing that *EVA* is a poor performance measure. It may well be the case that *EVA* is less distorted than most, if not all, alternative measures used in corporate practice. Our results only show that *EVA* is not first best, as it can be manipulated. Because this is likely to hold for all performance measures that can be realistically used in practice,[3] *EVA* may actually well be second best. We will return to this in the conclusion.

The contribution of this study is twofold. First, it empirically assesses the distortion of a widely used performance measure in the remuneration packages of executive board members. Second, the paper extends the Courty and Marschke test for distortions in performance measures to accommodate the potential endogeneity in the selection of performance measures (in this case,

---

[2] We will discuss the differences between *RI* and *EVA,* and the possible consequences of these differences for our analysis, in Section 3. Using *EVA* data obtained from Stern Stewart is not possible in our application (see Section 3).

3 In the words of Murphy (2012, p. 38) "Conceptually, the "perfect" performance measure for a CEO is the CEO's personal contribution to the value of the firm. Unfortunately […] the available measures will inevitably exclude ways that the CEO creates value, and include the effects of factors not due to the efforts of the CEO, or fail to reveal ways that the CEO destroys value. The challenge in designing incentive plans is to select performance measures that capture important aspects of the CEO's contributions to firm value, while recognizing that all performance measures are imperfect and create unintended side effects."

EVA), so that it can be used in non-experimental settings. This is important because it is rarely the case that economists can get data from a truly experimental setting. Thus, our difference-in-differences approach to their test is of general applicability to real-world incentive contracts.

2. DETECTING DISTORTIONS IN A PERFORMANCE MEASURE

The test developed by Courty and Marschke (CM) detects distortions by assessing the behavioral responses to changed incentives. It requires performance to be observed under (at least) two different compensation schemes, where the relative weight attached to the measure under study should have increased going from one scheme to the other. Let p_1 and p_2 therefore denote two different performance measures of the agent's efforts and let V reflect the value of the firm the principal is interested in. Assume that at some moment in time (denoted t_0) the principal changes the incentive weights on p_1 and p_2 in the agent's compensation contract from (β_1, β_2) to (β_1', β_2'). In particular, assume that the principal introduces p_2 (i.e. $\beta_2' > \beta_2 = 0$), instead of no measure at all, or an alternative performance measure p_1 used before (i.e. $\beta_1 \geq \beta_1' = 0$). As a result, the relative weight β_2/β_1 on performance measure two increases. This corresponds to the situation studied in our empirical analysis, where sample firms introduced *EVA* (=p_2) to replace their existing performance measure(s) (=p_1).

Courty and Marschke show that p_2 is distorted if the statistical association between p_2 and V decreases upon an increase in the relative weight β_2/β_1. Here, statistical association can be measured either by the correlation between p_2 and V, or by the slope coefficient of regressing V on p_2. If distortion is detected, it can be of two different kinds. First, the performance measure may induce the agent to exert (more) productive effort, but in the wrong quantities, because effort is imperfectly valued at the margin. Second, the agent may be motivated to undertake unproductive actions that increase the value of the performance measure and thus his reward, but are detrimental to firm value (*gaming*). Courty and Marschke show that the two types of distortion can be distinguished based on the direction of changes in the covariance between p_2 and V, but only if the additional assumption is made that overall incentives are not weakened over time. In that case, a decrease in the covariance between p_2 and V is conclusive evidence of gaming. Our empirical analysis therefore uses the following two test criteria developed by CM:

(CM.1) Performance measure p_2 is distorted if $Corr(p_2, V)$ decreases upon an increase in the relative weight (β_2/β_1) of this measure in the contract;

(CM.2) If $\beta_1^{/} + \beta_2^{/} \geq \beta_1 + \beta_2$, then a decrease in $Cov(p_2, V)$ upon an increase in (β_2/β_1) implies a gaming distortion in performance measure p_2.

The CM test is based on the assumption that changes in the levels of association between p_2 and V are solely caused by changes in the agent's effort choices. In their experimental setting, this assumption is likely to hold. In practice, however, there may be confounding factors.

To address this issue of confounding factors in our non-experimental setting, we compose control groups to benchmark the results of the firms that adopted EVA ($=p_2$). Firms are matched: each firm in the treatment group (using newly introduced EVA from time t_0 onwards as the (sole) performance measure in the board's compensation contract) is matched with a comparable firm in the control group (using neither EVA nor any similar measure in the time span studied before and after t_0). Comparing the changes in association in the treatment group to those in the control group allows us to identify changes caused by altered effort incentives (assuming that changes in other factors are similar for treatment and control firms). We thus use a difference-in-differences approach to the CM tests described above.

We set up two control samples. The first matches treatment firms to firms of the same size and industry class. This control group is able to mitigate the bias resulting from changes in (economic) circumstances. The second control group is set up to mitigate the endogeneity bias due to self-selection. The introduction of EVA (or any other performance measure) to reward management performance is a deliberate strategic decision—and not based on random assignment.

For example, firms that observe a temporary high correlation between *EVA* and firm value may be especially inclined to adopt *EVA* for reward purposes. Sheer regression to the mean then results in a lower correlation in the period after adoption—even when distortion plays no role at all. Using an (alternative) appropriate control group, our difference-in-differences approach addresses these confounding factors and can distinguish (self-selection-induced) regression to the mean from distortion.

3. EMPIRICAL APPLICATION TO ECONOMIC VALUE ADDED

This section discusses the residual income-based performance measure *EVA* and our empirical measure of firm value (that is, relative total shareholder return (*RTSR*)).

3.1 Residual Income and Economic Value Added

Economists have long acknowledged that an appropriate measure of value creation for firms should be based on the difference between earnings and the costs of capital employed (see Hamilton, 1777; Marshall, 1890; Biddle et al., 1997). In the twentieth century, various explicit measures of value creation were operationalized. These are called Residual Income (*RI*) based measures. They are often used as a performance indicator of (divisions or entire) firms and used as a basis for the reward of board members, because of their alignment with the main goal of for-profit organizations (that is, to add value to the owners' wealth in excess of their opportunity costs).

Residual income is defined as net operating profits after tax (*NOPAT*), minus a charge for invested capital.[4] This capital charge equals the firm's weighted average cost of capital (*WACC*), times the amount of capital employed. The *WACC* essentially weighs the minimally required return on equity and the required interest rate on debt according to the relative equity and debt levels in the financial structure of the company. Residual income is operationalized as follows:

$$RI = NOPAT - WACC \times Capital\ Employed. \quad (1)$$

[4] Here, operating profits are profits before deducting the after-tax costs of interest expenses to debt holders.

In the 1980s, Stern Stewart & Co. developed and trademarked a version of residual income, labeled Economic Value Added (*EVA*), which made some adjustments to *RI*, both standard and discretionary.[5] EVA became a very popular management tool for measuring company performance and a basis for the remuneration of board members. Many companies, such as Eli Lilly and Polaroid, started using *EVA* measures in their executive bonus schemes around the mid 1990s, the mean (median) year of introduction is 1995 (1996). Because of the fact that *EVA* was such a popular performance measure that is believed to be relatively undistorted, we decided to apply the CM test to companies who have adopted this measure in comparison to non-adopters.

There are two additional motivations for studying a sample of *EVA*-adopters. First, *EVA* adopters predominantly use *EVA* as the sole measure in the board's (short-term) incentive plan, due to the strong recommendation by Stern Stewart to do so (Stewart, 1991). The sole use of *EVA* is necessary given the assumptions of the model and the tractability of their validity, see Section 2. Second, adopting *EVA* as a RI based measure in the compensation plan is a deliberate, radical and significant change, supervised by a professional company, i.e., Stern Stewart. This radical and significant shift is also vital given the difference-in-differences approach we adopt.

Given these attractive features of the sample of EVA –adopters for our current purpose, we prefer the study of that group compared to the wider group of all companies that adopt an RI-based measure of any other kind in ways that are less uniform. Nevertheless, the focus of our empirical analysis is the accounting-based value of *EVA* that omits the (standard and discretionary) adjustments, see below. Thus, we effectively study the effect of adopting *RI* as a performance

---

[5] Biddle et al. (1997, p. 306) provide a long list of examples of these adjustments, including the capitalization and amortization of research and development and certain marketing costs, adding the change in bad debt allowances and subtracting marketable securities and construction in progress. See also Young (1999) for an elaborate discussion.

measure in the CEO's compensation contract for the selection of firms that claim to have started using a particular form of it, i.e. the trademarked version named *EVA*.

The early average adoption year, together with the required long time series of data before (and after) the introduction of *EVA* and the unavailability of Stern Stewart adjustments earlier than 1990, compels us to apply the accounting-based value of *EVA* as defined in equation (1) and to refrain from the Stern Stewart adjustments in our empirical tests. We acknowledge that the adjusted value of *EVA* may differ from the accounting-based *EVA* calculation in equation (1). Yet we expect these differences to be of relatively minor importance for our analysis. First, in practice the typical *EVA*-adopter only makes a few of these adjustments, if at all (see Malmi and Ikaheimo, 2003; Martin and Perry, 2000; Young, 1999; Young and O'Byrne. 2001). Second, as Biddle et al. (1997, p. 314) show, the correlation between *EVA* based on the published numbers by Stern Stewart and the accounting-based value of *EVA* amounts to 0.90. Anderson et al (2005) similarly obtain a strong relationship between unadjusted and adjusted *EVA*. Third, Wallace (1997, section 4.6.2) finds that the effect of an *EVA*-based compensation plan on managers' behavior is independent of these adjustments to earnings.[6]

We measure the three ingredients of accounting based *EVA* in equation (1) as follows. *NOPAT* is the operating profit that accrues to shareholders and debt holders, and was obtained from the Thompson Worldscope database in the following two ways:

$$NOPAT_a = Operating\ Profit \times (1 - Tax\ Rate) \quad (2a)$$

$$NOPAT_b = Operating\ Profit - Taxes\ paid\ (received) \quad (2b)$$

[6] Our approach of ignoring *EVA*-adjustments would potentially be problematic if these adjustments varied systematically across companies (with characteristics that go together) with different levels of *RTSR*.. We are not aware of any empirical evidence supporting (or undermining) this case.

While in theory both approaches would logically yield the same result, they are, in practice, slightly distinct.[7] This study takes the average of the two and reports the results obtained when using either one of these definitions on its own in a separate robustness check in online Appendix B.

The weighted average cost of capital (*WACC*) is determined by the investors' required rate of return, which is not directly observable. Various judgmental decisions and arbitrary choices are involved when calculating the *WACC*, possibly leading to large variations.[8] Therefore, remuneration committees of virtually all companies use predetermined or 'pre-calculated' *WACC* levels. Following their practice, we adopt a commonly used flat *WACC* level of 10 percent. In our robustness check we include the other commonly used WACC levels of 12 and 8 percent in determining *EVA* (Stark and Thomas, 1998).[9]

The amount of *Capital Employed* is proxied by the average annual book value of total assets employed by the company, based on accounting data taken from the WorldScope database. The average is calculated by taking the average of the *year end* and *year start* values.

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

---

7 For example, negative tax rates are not revealed in the Thompson WorldScope database, thus rendering the reimbursement of taxes impossible if using equation (2a)—but not if using equation (2b).

8 The calculated *WACC* level is affected by judgmental decisions regarding the length and periodicity of the periods included in the calculations of β, a component of the capital asset-pricing model used to determine the required rate of return on equity, the levels of the risk-free rate and the risk spread. Moreover, required interest rates do not unambiguously follow from the observed annual interest payments, dependent on a company's possible suspended or accelerated interest payments.

9 We note that the implicit assumption that a unique WACC level applies to all company-year observations in the sample is unrealistic. Even stronger, WACC levels are even unlikely to be the same for distinct divisions/business units within companies (Kruger, Landier and Thesmar, 2011). Therefore, it would be reasonable to employ differential predetermined WACC-levels for various company-year observations, for instance dependent on the risk associated with each company in each year. However, it is difficult to think of uncomplicated observable characteristics that affect the variation in the WACC level across company-year observations and the literature appears relatively silent about the existence of such indicators. As long as the variation in WACC levels does not vary systematically between treatment and control companies, the results of our difference in difference approach are likely to remain unaffected.

not determine bonus levels. We therefore take as the performance measure actually studied: $\Delta EVA = EVA_t - EVA_{t-1}$, with EVA_t calculated on the basis of equation (1).

3.2 Relative Total Shareholder Return (*RTSR*) as a measure of firm value

The most widely accepted goal of (for-profit) organizations is maximization of value creation (see Baker 1992; Kaplan & Norton, 1996). In companies, value creation is commonly 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. This study uses Relative Total Shareholder Return (*RTSR*) as the relevant measure for the firms' objective function, for three reasons. First, contrary to *MVA*, *(R)TSR* includes dividend payments to shareholders, a potentially important source of shareholder returns. Second, absolute measures of value creation such as *MVA* are sensitive to differences in company size, whereas *(R)TSR* is not. Third, using *RTSR* instead of *TSR* allows us to control for variations in economic conditions in the different years of the adoption of *EVA*.

4. CHARACTERISTICS OF *EVA*-ADOPTERS AND NON-ADOPTERS

The group of 67 treated firms consists of basically all companies listed on the New York Stock Exchange that publicly communicated adoption of ΔEVA in the compensation scheme of the company's board members (before 2004) and whose basic accounting and market information could be found. Appendix A provides the details of our data search method and includes a list of all treatment firms, including the year in which ΔEVA was implemented (mean 1995, median 1996). Seventy-one percent of the companies in the treatment group use ΔEVA as the sole measure in the (short-term) incentive plan, as recommended by Stern Stewart (Stewart, 1991), whereas the average weight attached to *EVA* is 84 percent.[10]

Throughout the study, we consider two alternative time spans k ($k=8$ and $k=10$) when calculating the relevant statistics over period T=I $\{t_0\text{-}k,..,t_0\text{-}1\}$ before the introduction of performance measure $p_2 = \Delta EVA$, and period T=II, $\{t_0+1,..,t_0+k\}$ afterwards.[11] The determination of t_0 is discussed in Appendix A. All statistics are calculated after removing outliers from the sample, defined as a company-year observation where either *RTSR* or ΔEVA is in the extreme 1.5 percentile of their distributions.

To each of the 67 treated firms, we matched a NYSE-listed control company that did not use an *RI*-based performance measure, neither in T=I nor in T=II. Following Wallace (1997), matching takes place based on four-digit *SIC* codes and company size, see Appendix A and Table A1. Our analyses require at least four company-year observations both before t_0 and after t_0 of all the variables in the analyses (three observations in terms of changes). This reduces the sample to 40 treatment and 49 control firms and to 34 *pairs* matched on size and sector to perform all of the analyses.[12]

As discussed, the Courty and Marschke (CM) approach suffers from a potential endogeneity bias. If the decision to use *EVA* for rewarding management performance is partly driven by the statistical measures that appear in the CM tests, these tests are potentially confounded by regression to the mean. A priori this is likely to be the case because (theoretically) the two most probable drivers of adopting a particular performance measure are (low) noise and (little) distortion. Noise is typically measured by the (time-series) variance in the performance measure due to factors

---

[10] For the firms that did not use ΔEVA as their sole measure, "implementation of ΔEVA" is defined as an increase of the weighting of ΔEVA to at least 10%, where ΔEVA was not used before.

[11] Lower levels of k would generate a rather small number of observations for the required regressions and rather imprecise pre- and post-treatment statistics based on these time series. Higher k's are difficult, due to data limitations.

[12] Paired t-tests show that the treatment and control group are not significantly different in terms of ΔEVA, *RTSR*, size, company age since IPO before the year t_0, (for $k=8$ or $k=10$). Further inspection of the data shows that *EVA* adoption is not confounded or driven by CEO changes around the date of activation of the measure.

beyond the agent's control (see Prendergast, 2002; Coles et al., 2006), whereas a commonly used (but misguided) inverse measure of distortion is the correlation between the performance measure and company value (see Biddle et al., 1997; Stark and Thomas, 1998; Feltham et al., 2004).[13] Thus, one might expect that —before treatment— the average variance of Δ*EVA* is lower whereas the correlation between Δ*EVA* and firm value is higher for firms in the treatment group vis-à-vis firms in the control group.

Table 1: Mean values before treatment (i.e. before t_0) of *Corr*(Δ*EVA,RTSR*), *Std*(Δ*EVA*), and *Cov*(Δ*EVA,RTSR*) in treatment and control samples: Matching based on SIC/Size and Correlation

			Mean value before treatment (T=I)		
	matching	**# years**	*Treatment*	*Control*	*Difference [p-value]*
Panel A: *Corr(ΔEVA,RTSR)*	**SIC/Size** (n=34)	*k=8*	.295	.012	.282** [.016]
		k=10	.272	.057	.215** [.036]
	Correlation (n=31)	*k=8*	.232	.235	-.003 [.891]
		k=10	.260	.263	-.003 [.730]
Panel B: *Std(ΔEVA)*	**SIC/Size** (n=34)	*k=8*	34.083	40.694	-6.611 [.270]
		k=10	35.925	39.069	-3.144 [.380]
	Correlation (n=31)	*k=8*	29.076	35.153	-6.077 [.639]
		k=10	32.311	38.658	-6.346 [.594]
Panel C: *Cov(ΔEVA,RTSR)*	**SIC/Size** (n=34)	*k=8*	1.707	1.042	0.665 [.336]
		k=10	1.307	1.005	0.301 [.405]
	Correlation (n=31)	*k=8*	.726	2.737	-2.011 [.102]
		k=10	.761	1.711	-.950 [.322]

The columns labeled 'Difference' report the difference between the average values (of the correlation between Δ*EVA* and *RTSR*, the standard deviation of Δ*EVA* and the covariance between Δ*EVA* and *RTSR,* respectively) in the treatment and control group. Within squared brackets appear the p-values of paired t-tests that compare the means across groups (on a pair-wise basis). A significant difference at the 10% (5%) [1%] level is denoted by * (**) [***]. The SIC/Size matching tests are one-sided, the alternative hypothesis being that treatment firms have higher (pre-treatment) values of the correlation and the covariance, and lower values of the standard deviation. For the correlation, matched sample tests are double-sided.

[13] See, for example, Baker (2002, p. 736): "… 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) shares Baker's criticism.

The upper halves of Table 1 show that the treatment group and (first) control group indeed differ as expected, due to self-selection in terms of the before-treatment values of the three key CM statistics. The *one-sided* pairwise t-tests in the last column of Table 1 show that the before-treatment values of the correlation (Panel A) and covariance (Panel C) are indeed higher in the treatment group than in the control group, whereas the standard deviation of ΔEVA is lower in the treatment sample (Panel B). The only significant difference is the correlation between ΔEVA and *RTSR* (Panel A).[14]

A multivariate (probit) analysis that studies the drivers of selection into the treatment sample supports the picture communicated in Table 1. Here, the dependent variable is a dummy equal to one for treatment firms and zero for control firms, whereas the set of independent variables includes the time-series correlation between ΔEVA and *RTSR* and the time-series standard deviation of ΔEVA. Also in this analysis, the only observed significant determinant of selection into the treatment sample (given a similar control sample in terms of sector, company size and age) is the correlation between ΔEVA and *RTSR* (before treatment). The higher this correlation is, the more likely that the performance measure is adopted. Given this, a decreasing correlation upon adopting the measure in the board's reward contract 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 Appendix A for which we have enough information).[15] Firms in this group are matched pairwise to firms in the treatment group on the sole basis of the pre-treatment correlation between ΔEVA and *RTSR*, separately for $k=8$ and $k=10$.[16] This k-specific matching serves as a guarantee that each pair consists of a treatment- and a control firm whose pre-treatment correlations between ΔEVA and *RTSR* do not differ by more than 0.1. The resulting correlation-based matched samples (for $k=8$

[14] Table B1 in online Appendix B shows that these results are similar when using alternative measures of ΔEVA, based on alternative definitions of *NOPAT* or alternative *WACC* levels.

[15] Observations omitted from the first SIC/Size matched sample (treatment or control) due to missing data for the counterpart (in terms of SIC/Size), are reconsidered for inclusion in the second (pre-correlation matches) sample.

[16] Exact propensity score matching is unnecessary, as matching is based on one explanatory variable.

and $k=10$) are somewhat limited in size due to the different distributions in the two samples of the pre-treatment correlations. We must omit firms with very high pre-treatment correlations from the treatment sample and firms with very low correlations from the control group. The resulting sample size is 31 matched pairs for both $k=8$ and $k=10$.[17]

The lower halves of Table 1 show that the differences between the treatment- and this matched control group are indeed negligible in the pre-treatment period t_0-k to t_0-1 in terms of the key CM statistics for both $k=8$ and $k=10$.[18] Therefore, we think it is useful to present in Section 5 the results of the CM tests based on a comparison of both the first (SIC/Size-based) and the second (correlation-based) set of treatment- and control firm pairs.

5. EVIDENCE OF DISTORTIONS AND GAMING

This section tests for the presence of distortions in ΔEVA, using tests (CM.1) and (CM.2) from Section 2.[19] Table 2 reports the correlation levels between ΔEVA and $RTSR$ before and after activation of the performance measure for both sets of treatment-control samples and for both time spans k. The right-hand side reports the difference between the correlations after and before treatment ($\Delta Corr$) and the difference-in-differences between the treatment and the control group ($\Delta^2 Corr=[\Delta Corr]_{Treatment}-[\Delta Corr]_{Control}$). Hence, this last column is the one that indicates whether or not there is evidence of distortion according to test (CM.1).

---

17 Paired t-tests show that this matched control group is also not significantly different from the treatment group in terms of ΔEVA, $RTSR$, size, company age since IPO before the year t_0., (for $k=8$ or $k=10$). Further inspection of the data shows that EVA adoption is not confounded or driven by CEO changes around the date of activation of the measure.

18 This test is two-sided, given the alternative hypothesis that the two samples have *different* average values of the pre-treatment correlations (unlike the expectations for the SIC/Size matched sample), without prior expectations about the sign of the difference.

19 These tests are also performed for the alternative measures of ΔEVA based on different calculations of $NOPAT$ and different $WACC$ levels, with the results reported in online Appendix B. In all cases the findings for the alternative measurements are comparable to those reported here.

Table 2: Testing for the degradation of the correlation coefficient *Corr(ΔEVA,RTSR)* (test (CM.1))

		Treatment			*Control*			*DID*
matching	**# years**	*before t_0*	*after t_0*	*ΔCorr [p-value]*	*before t_0*	*after t_0*	*ΔCorr [p-value]*	*Δ^2Corr [p-value]*
SIC/Size (n=34)	*k=8*	.295	.070	-.225*** [0.005]	.012	.104	0.092 [0.855]	-.317*** [0.006]
	k=10	.272	.080	-.193*** [0.007]	.057	.106	0.049 [0.694]	-.241** [0.025]
Correlation (n=31)	*k=8*	.232	.006	-.226** [0.012]	.236	.173	-.062 [0.206]	-.164** [0.029]
	k=10	.260	.049	-.211*** [0.010]	.263	.144	-.119* [0.077]	-.093 [0.135]

The p-values reported in square brackets are based on a one-sided paired t-test, and confirm the expectation that the decrease of the correlation levels is larger in the treatment group than in the group of control firms. A significant difference at the 10% (5%) [1%] level is denoted by * (**) [***]. *DID* refers to difference-in-differences.

The results support distortion of the performance measure Δ*EVA*. The correlation between Δ*EVA* and *RTSR* decreases significantly in the treatment group, but (typically) not in the matched control group. The difference-in-differences are large and mostly significant. They are smaller, however, for the correlation-based treatment-control comparisons than for the SIC/Size-based treatment-control comparisons. This indicates that regression to the mean explains part of the decreased correlation between Δ*EVA* and *RTSR* once Δ*EVA* is implemented. Yet regression to the mean is certainly not the single driving force, as is indicated by the significantly negative estimates of Δ^2*Corr* for the correlation-matched sample of pairs (when *k=8*). Degradation plays an important role as well. Overall, Table 2 thus provides plausible evidence for the distortion of Δ*EVA* as a performance measure.

In online Appendix C we investigate the changes in the slope coefficient from regressing *RTSR* on Δ*EVA* as an alternative test of distortion and obtain exactly the same conclusion. The evidence in favor of distortion is thus robust.

Distortion may either point at suboptimal amounts of productive effort, or at unproductive gaming activities detrimental to firm value. Evidence of whether gaming plays a role is obtained by applying test (CM.2.) that measures the change in the covariance. The results reported in Table 3 display a clear pattern in support of gaming: the reductions in the covariances in the treatment group are significantly greater than the comparable reductions in the control group. This holds for both matching methods and for both levels of *k*.

Table 3: Testing for the degradation of the covariance *Cov(ΔEVA,RTSR)* (test (CM.2))

		Treatment			*Control*			*DID*
matching	**# years**	*before* t_0	*after* t_0	*ΔCov* *[p-value]*	*before* t_0	*after* t_0	*ΔCov* *[p-value]*	Δ^2*Cov* *[p-value]*
SIC/Size (n=34)	*k=8*	1.707	0.929	-0.778 [0.289]	1.042	4.293	3.251## [0.981]	-4.029** [0.016]
	k=10	1.307	0.675	-0.632 [0.296]	1.005	4.997	3.992### [0.991]	-4.623** [0.016]
Correlation (n=31)	*k=8*	0.726	-0.148	-0.874 [0. 254]	2.737	5.789	3.051## [0.968]	-3.926** [0.042]
	k=10	0.761	-0.473	- 1.234* [0.091]	1.711	5.752	4.041## [0.989]	-5.276*** [0.006]

The p-values reported in square brackets are based on a one-sided paired t-test, and confirm the expectation that the decrease of the covariance levels is larger in the treatment group than in the group of control firms. A significant difference at the 10% (5%) [1%] level is denoted by * (**) [***]. Significant differences opposite to the expected direction are denoted by # (##)[###]. *DID* refers to difference-in-differences.

Using a decrease in the covariance to detect gaming assumes that incentives are not weakened over time (see test (CM.2) in Section 2). This assumption is likely to be met for our sample. Murphy (2012) shows strong evidence of the increased incentive intensity for US based listed firms between 1992-2011. His Figures 2.4 and 2.5 reveal that the fraction of variable pay in overall pay has significantly increased over time..[20] We are thus fairly confident that this holds for our sample as well**.**

[20] Measuring incentive intensities is not straightforward and the various measures that are commonly used all point in the same direction: the incentive intensity of CEOs has increased across the board in the period of our study. CEOs effective ownership of stock as % of total number of shares outstanding (i.e. Jensen and Murphy (1990)'s pay performance sensitivity measure) steadily increased from 1992 to 2004, just as CEOs "equity at stake" did, i.e. CEOs change in wealth for a 1% change in the value of the firm, see Figures 2.8 and 2.9 in Murphy (2012). Frydman and Saks (2010, Figure 6) also observe a strong and steady increase in both Hall and Liebman (1998)'s equity at stake and Jensen and Murphy's pay performance sensitivity from 1980 to 2005.

All in all, we conclude that the performance measure ΔEVA is distorted. Our results suggest that part of this distortion is due to gaming of the performance measure by board members. The changes in statistical association between ΔEVA and firm value upon incorporating ΔEVA in the board's remuneration plan are also driven by regression to the mean. Correcting the CM tests for self-selection is thus imperative.

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 Economic Value Added in particular—have been subject to such evaluations, leading to mixed results (e..g Biddle et al., 1997; Chen and Dodd, 1997; Feltham et al., 2004; O'Byrne, 1997; Stark and Thomas, 1998; Wallace 1997). More recently, however, Baker (2002) has shown that what matters is not the association between the *levels* of the performance measure and the company's objective function, 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 distortion is harder to assess. Courty and Marschke (2008) recently developed a suitable empirical test to detect distortion in performance measures. This paper applies their test to the (accounting based value of the) performance measure EVA, a Residual Income-based performance measure widely used in corporate practice. The test is adapted in order to cope with self-selection and timing decisions. In particular, we take a difference-in-differences approach, where the control group consists of a matched sample of suitably selected firms.

Overall, our test results indicate that the accounting based value of EVA is a distorted performance measure. Differences (in differences) are somewhat more pronounced when the treatment-control group pairs consist of firms that are similar in terms of industry and size than when pairs are formed based on similar pre-treatment correlations between $RTSR$ and ΔEVA (the main driver of self-selection into the treatment). This indicates that regression to the mean explains part of the decrease in correlation between *RTSR* and Δ*EVA* after the activation of *EVA*. Yet the significant differences that we observe for our second control group reveal that degradation plays a key role as well. Moreover, under the reasonable assumption that overall incentives have not weakened over time (cf. Murphy, 2012, Frydman and Saks, 2010), our finding that the covariance between *RTSR* and Δ*EVA* decreases more for the treatment firms supports gaming. There is, thus, reasonable evidence that the distortions detected are partly due to gaming.

Our finding that *EVA* is a distorted performance begs the question exactly *how* it can be gamed. The empirical accounting literature tentatively suggests that managers may engage in 'earnings management' and 'short-termism' to artificially boost current *EVA* at the expense of future *EVA*. First, using *EVA* may give managers incentives to manipulate when actual accounting profits are reported, as to influence pay that depend on reported earnings (Healy, 1985; Young and O'Byrne, 2001).[21] Second, distortions may arise from *EVA* being a short-term, single-period performance measure. Managers may want to avoid negative *EVA* projects—even if these projects are profitable in the long run (see Bromwich and Walker, 1998; O'Hanlon and Peasnell, 1998). In line with this, Wallace (1997) finds suggestive evidence that managers whose rewards are *EVA*-based may actually under-invest in projects that render a positive net present value relative to managers whose rewards are based on more traditional (earnings-based) performance measures. Relative to non-adopters, *EVA*-adopters dispose of more assets and decrease their new investments, whereas they use their remaining assets more intensively.[22]

A potential interesting avenue for future research would be to complement our general

[21] The incentive to manage earnings is provoked by the use of upper limits to restrict bonus payments, although this practice counters the recommendations of Stern Stewart & Co.

22 As noted by Wallace (1997, p. 287), it is difficult to establish whether the observed changes in behavior are value-decreasing. Rogerson (1997) and Reichelstein (1997) show that, theoretically, *EVA*-based compensation can provide first-best investment incentives if depreciation schedules are chosen such that investment costs are spread out over the investment's lifetime exactly proportionally to the benefits. The intuition is that *EVA* then reflects the value created by the manager at any given point in time.

test based on Courty and Marschke (2008) with an analysis of accounting information to test for some of the specific ways of gaming discussed above. If specific gaming activities could be empirically identified for our sample of firms, this would reinforce our conclusions. At the same time, there are likely to be numerous ways in which a performance measure can be manipulated, and managers may be very creative in findings the more opaque ways to do so. The general test of Courty and Marschke has the advantage that distortions can be detected without being omniscient of all the possible ways in which a measure can be gamed.

As noted in the Introduction, our findings do not imply that EVA is a poor performance measure. First, our empirical analysis has little to say about the economic relevance of the distortion identified in terms of efficiency losses. This would require observing the counterfactual, i.e. changes in firm value under fist best incentives. Second, our analysis is unable to make comparisons of the distortions and thus the relative welfare losses when any other performance measure is used. Another interesting topic for future research would therefore be to extend the Courty and Marschke (2008) approach to find ways in which the distortions detected in different performance measures could be meaningfully compared. In the end it may be more valuable to know which out of several available measures is least subject to gaming (in terms of efficiency losses), than just establishing that a given measure can be gamed.

Appendix A Data

RTSR

The *RTSR* of firm *j* in year *t* ($RTSR_{j,t}$) is obtained by normalizing the *TSR* of this firm in year *t* with the *TSR* of the S&P500 index in that year:

$$RTSR_{j,t} = \frac{I_{j,t} / I_{j,t-1}}{I_{m,t} / I_{m,t-1}} - 1$$

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, dividend payouts, the effects of stock splits, stock shares and so forth.[23] Subscript *m* refers to the market, *j* refers to a firm and *t* is a year index. The fiscal year that includes the month in which the performance measure was implemented in the (matched) treatment firm is recorded as the year t_0.

Selection of treatment firms

To identify the population of ΔEVA-adopters, we used three sources: (i) a list composed by Stern Stewart & Co., taken from their company report: "The comparative stock market performance of Stern Stewart Clients", (ii) the sample of treatment firms in Wallace (1997),[24] and (iii) a search through the encompassing EDGAR database of the US Securities and Exchange Commission using keywords "*EVA*" and "Economic Value Added".[25] The combined lists provided 149 companies that adopted ΔEVA for performance measurement. An analysis of the individual proxy statements of these firms revealed that 74 of them explicitly adopted ΔEVA as a basis for

---

[23] In accordance with general practice, stock indices are evaluated over a period of three months, which smoothes the influence of the day-by-day volatility of the stock market.

[24] Wallace (1997) focuses on residual income-based compensation plans. We selected the EVA users from his sample based on the assessment of companies' proxy statements.

[25] EDGAR, the Electronic Data Gathering, Analysis and Retrieval system, performs automated collection, validation, indexing, acceptance, and forwarding of submissions by companies that are required by law to file forms with the Securities and Exchange Commission.

board member remuneration.[26] For seven of these companies none of the crucial accounting or market data was available at all. This resulted in 67 treatment firms (see Table A1 below).

Determining the treatment start and end

The exact implementation year t_0 for each treatment firm was deduced from their proxy statements. Most of the firms explicitly mentioned 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." (Becton Dickinson & Co, DEF14A statement Fiscal Year 2007).

We also checked if and when *EVA*-adopters abandoned the use of this measure as part of the compensation contract later on. Roughly 58% did so in the 11 years (at most) that we observed after the measure's introduction—on average, after 7.5 years. Thus, only relatively few company-year observations are involved (approximately 7%). These observations are set to missing values in the regressions for the relevant treatment-control pairs. The results are qualitatively similar when we ignore the elimination of the performance measure.

Selection of matched control firms in the first (SIC/Size) control sample

As in Wallace (1997), firms are first matched on the 4-digit *SIC* code. Next, the firm that is closest in size is selected as control. Size is measured by sales volume in the year prior to the treatment firm's adoption of ΔEVA (i.e. year t_0-1). 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.[27] In these cases, matching takes place on 3-digit, or in a few cases even two-digit *SIC* codes. Two firms were used twice as a control firm. We treat these as independent observations in the analyses.

[26] Kleiman (1999) followed a similar search procedure with slightly different criteria and obtained a set of 71 *EVA*-adopters.

[27] 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.

Table A1: Companies in the treatment- and control group

Treatment company	Activation year	SIC	Paired control-group company
Coca Cola Company	1983	2086	Pepsico Inc
Grainger WW Inc	1989	5000	Allied Waste Industries Inc
Briggs & Stratton Corp.	1990	3510	Caterpillar
Crane Company	1990	3490	Parker-Hannifin Corp.
Ball Corp.	1991	3411	Crown Holdings Inc
Manitowoc Company Inc	1993	3531	Alico Inc
Milacron Inc	1993	3559	Wellco Enterprises Inc
Acxiom Corp.	1994	7374	Total System Services Inc
Centura Banks Inc	1994	6022	Doral Financial
Donnelley RR & Sons Company	1994	2754	Interpublic Group Companies Inc
Fleming Cos Inc	1994	5141	Performance Food Group Company
Georgia Pacific Corp.	1994	2400	Glatfelter
Jarden Corp.	1994	4783	American Greetings Corp.
Maritrans Partners	1994	4499	Sea Containers Limited
PerkinElmer Inc	1994	3841	Applera Corp.
Ablest Inc	1995	7361	National Technical Systems Inc
Aquila Inc	1995	4922	El Paso Corp.
Armstrong Holdings Inc	1995	3996	Interface Inc
Becton Dickinson & Company	1995	3841	3M Company
ELI Lilly & Company	1995	2834	Abbott Laboratories Inc
Great Plains Energy Inc	1995	4911	Ohio Power Company
Guidant Corp.	1995	3845	Fischer Imaging Corp.
Inacom Corp.	1995	5045	En Pointe Technology Inc
Lyondell Chemical Company	1995	2869	Dow Chemicals Company
Officemax Incorporated	1995	5943	Staples Inc
Tektronix Inc	1995	3825	COHU Inc
ADC Telecommunications Inc	1996	3661	Arris Group Inc
Avery Dennison Corp.	1996	2891	Cytec Industries
Bausch & Lomb Inc	1996	3851	Oakley Inc
Bowater Inc	1996	2621	Glatfelter
GC Cos Inc	1996	7832	Carmike Cinemas Inc
Johnson Outdoors Inc	1996	3792	Fleetwood Enterprises
K-Swiss Inc	1996	5139	Kenneth Cole Productions Inc
Kimball International Inc	1996	2517	Ethan Allen Interiors Corp
Magnetek Inc	1996	3679	BTU International
Marathon Oil Corp.	1996	2911	Hess Corp.
Millennium Chemicals Inc	1996	2899	PPG Industries Inc
Miller (Herman) Inc	1996	2522	Flexsteel Industries
Murphy Oil Corp.	1996	2911	Frontier Oil Corp. Commerce
Olin Corp.	1996	3341	Rockwood Holdings Inc
Polaroid Corp.	1996	3827	Cooper Companies Inc
SPX Corp.	1996	3823	Ametek Inc
SVB Financial Group	1996	6021	Alabama National BanCorporation
Tupperware Brands Corp.	1996	3089	Aptargroup Inc
Vulcan Materials Corp.	1996	1422	MDU Resources Group Inc
Allied Holdings Inc	1997	4213	Covenant Transport Inc
Columbus McKinnon Corp.	1997	3536	Alamo Group Inc
Federal Mogul Corp.	1997	3713	Dana Corp.
Material Sciences Corp.	1997	3479	Lindsay Corporation
MDI Inc	1997	3829	Mechanical Technology Inc
Noble Corp.	1997	1381	Abraxas Petroleum Corp
Penney JC Company Inc	1997	5311	Federated Department Stores Inc
Ryder System Inc	1997	7513	Amerco
Tenet Healthcare Corp.	1997	8062	Universal Health Services Inc
Tenneco Inc	1997	3714	BorgWarner Inc
Touch America Holdings Inc	1997	4899	Ohio Power Company
Webster Financial Corp.	1997	6021	Downey Financial Corp.
Whirlpool Corp.	1997	3630	Centex Corp.
Wolohan Lumber CO.	1997	5211	Building Materials Holdings Corp.
Great Lakes Chem. Corp.	1998	2819	Church & Dwight
ITT Corporation	1998	3812	General Dynamics
Servicemaster Company	1998	0782	Central Garden & PET Company
Standard Motor Products Inc	1998	3694	Motorcar Parts Of America
Worthington Industries Inc	1998	3312	AK Steel holdings
Best Buy Company Inc	1999	5731	Circuit City Inc
Genesco Inc	1999	5661	Brown Shoe Inc
Int Multifood Corporation	2001	5149	McCormick & Company Inc

References

Anderson, A.M., R.P.Bey, and A.C.Weaver. 2005. "Economic Value Added Adjustments: Much to Do about Nothing ?", Working Paper, Lehigh University.

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

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

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.

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

Chen, S. and J. 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, 2008. "A General Test for Distortions in Performance Measures", *Review of Economics and Statistics* 90(3), 428-441.

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

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 —Revisiting Previous Evidence", *Journal of Applied Corporate Finance* 16, 83-88.

Frydman, C. and R.E. Saks, 2010. "Executive Compensation: A New View from a Long Term Perspective, 1936-2005", *The Review of Financial Studies* 23, 2100-2138.

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

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

Healy, P., 1985. "The Effect of Bonus Schemes on Accounting Decisions", *Journal of Accounting and Economics* 7, 85-107.

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.

Jensen, M.C. and K.J. Murphy, 1990. "Performance Pay and Top Managerment Incentives", *Journal of Political Economy* 98, 225-264.

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

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

Kruger, P., A. Landier and D. Thesmar, 2011. "The WACC Fallacy: The real effects of using a unique discount rate", Working Paper, HEC Paris.

Malmi, T. and S. Ikaheimo, 2003, "Value Based Management Practices – Some Evidence from the Field", *Management Accounting Research* 14, 235–254.

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

Martin, J.D. and J.W. Petty, 2000, *Value Based Management: The Corporate Response to the Shareholder Revolution*, Harvard Business School Press, Boston, MA.

Murphy, K.J, 2012, "Executive Compensation: Where We Are, and How We Got There", forthcoming in: Constantinides, G., Harris, M. and R. Stulz, eds., *Handbook of the Economics of Finance*

O'Byrne, S., 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.

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.

Rogerson, W., 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.

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, 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 What You Pay for?" *Journal of Accounting and Economics,* 24, 275-300.

Young, S.D. and S. O'Byrne, 2001, *EVA and Value-Based Management*, McGraw-Hill, New York.

Young, S.D. , 1999. "Some Reflections on Accounting Adjustments and Economic Value Added", *Journal of Financial Statement Analysis* 4, 7-13.

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