The Balanced Scorecard: Suggestions for Rebalancing
Edward J. Lusk, Michael Halperin, BaoDong Zhang

Abstract

Recently, evidence has been presented that, in use, the Balanced Scorecard (BSC) is not realizing its potential and thus is in need of refocusing. We offer two BSC enrichments to address the needed rebalancing: (1) a direct connection to shareholder wealth through specific financial market measures to address the critique that the BSC is too heavily focused on internal financial measures of performance and so lacks a connection to shareholder wealth, and (2) the incorporation of a stakeholder context to address the issue that the BSC needs to widen its scope to consider the societal context within which the organisation operates. We rationalize and illustrate these suggestions with a study of the market effects identified for selected KLD issue-screens that seem to be ideal market drivers. We used as our market measures: Jensen’s $\alpha$, The Sharpe Performance Index, The Treynor Performance Index, and comparisons relative to The CRSP Standard Deviation Peer Group and the CRSP $\beta$ Peer Group. We find for all five market measures that organisations scored by KLD as being socially responsible outperformed organisations scored by KLD as having a social responsibility profile that raised concerns. Further, in terms of relative risk, these socially responsible organisations did not have a higher risk profile as measured by $\beta$.

Key words: Balanced Scorecard, KLD Issue-Screens, Market Drivers.

Introduction

Kaplan & Norton’s Balanced Scorecard (BSC) debuted in the Harvard Business Review in 1992 (Kaplan & Norton, 1992, 1993). It described a way for management to improve the organization’s competitive advantage by broadening the scope of evaluation. In addition to the usual Financial perspective, a lagged performance measure, the BSC also targets in its evaluation, the following leading aspects of financial performance: the organization’s Customer base, the constitution and functioning of the firm’s Internal Business processes and the necessity of Innovation and Learning as a condition for growth. Kaplan & Norton (1992, p. 78) clearly focus the BSC on financial dimensions of organizational performance. They note: “A failure to convert improved operational performance, as measured in the scorecard, into financial performance should send executives back to their drawing boards to rethink the company’s strategy or its implementation plans”. Over the years, the four constituent elements of the BSC have remained largely unchanged, with the exception of a modification in 1996 when Innovation and Learning was changed to Learning and Growth (Kaplan & Norton, 1996, 1996a). About ten years after its introduction, Kaplan & Norton (2001, p. 87) give the following summary view of the BSC: “Why has the Balanced Scorecard concept been so widely adopted by manufacturing and service companies, nonprofit organizations, and government entities around the world since its introduction in 1992?” In answering, they continue: “...previous systems that incorporated nonfinancial measurements used ad hoc collections of such measures, more like checklists of measures for managers to keep track of and improve than a comprehensive system of linked measurements”. Here it is important to point out that the nonfinancial measures they note are in reality the three leading aspects of financial performance mentioned above, and not the nonfinancial measures that one usually thinks of in the stakeholder context. At about the same time that the BSC was unveiled in the Harvard Business Review, three other events were underway that would eventually pressure the BSC to change. Consider now these three events and the changes that they create in the competition for capital.

1 We wish to express our appreciation for the helpful suggestions made by Professors Aronsson of the Umea University, Umea, Sweden, Alfred Luhmer, of the OVG University, Magdeburg, Germany and to Ms. Ellen Slack of the Lippincott Library of the Wharton School of the University of Pennsylvania, Philadelphia, PA, USA for her invaluable editorial assistance.
The New Kid on the Global-Block

The first event is the take-off of the Chinese economy after Deng Xiaoping’s dramatic opening of the Chinese economic-door to the west, reversing the “closed-door” policy necessary during the Mao-era re-birthing of China. The temptation for multinational corporations (MNC) to accept Beijing’s invitation to invest in China with its low salary base, highly skilled and enormous labour pool, and promised governmental stability was of course irresistible. This put pressure on all organizations to rethink global cost competition. It was all about protecting the bottom line and so being able to compete for funds in the debt and equity capital markets. The ensuing rush of MNCs to establish a “first-moved-in advantage” can only be described as a MNC-stampede to establish a presence in China (Spence, 2005). The significant flow of capital that has continued unabated for about 12 years has predictably put pressure on the Chinese monetary system. In 2004, China recorded such a high rate of growth, 9.5%, that its central planning committee is recom-mending steps to slow down foreign investment, to guard against the possibility of a destabilising inflation. According to the Wall Street Journal, 7 March 2005, “In his annual work report to the national legislature during the weekend, Mr. Wen (Premier Wen Jiabao) said the government is targeting fixed-asset growth of 16% -- still double the targeted growth rate for the overall economy, but sharply lower than last year's fixed-asset growth of 26%”.

The Electronic Looking-Glass

The second development, circa 1992, was the Internet. All of a sudden there was this new “gizmo” called the Internet that seemed a magical portal into the techno-future. Most in the usually conservative investment community were caught off guard. The promise of the “dot-coms” blinded the usually suspicious “show-me-the-cash-flow” investment bankers. Even banks, following the lead of the bedazzled venture capitalists, started dumping fund money into anything that ended in “.com”. Most models of investment evaluation were thrown out the window; caution and money were thrown to the wind. The prevailing sentiment in the mid-1990s seemed to be: We don’t know what these dotcoms are and we realise that they have no asset base to speak of, no possibility for profitability even in the intermediate future, and their cash flows look negative for the foreseeable future. All we know for sure is that their price will be much higher in a few months and we ask ourselves: What if we don’t get in right now? The possibility of such high expected returns was attracting a disproportionate share of risk-relative capital and so acted to pressure all organizations to improve their bottom lines so as to at least maintain their relative-risk positions. According to Higgins & Currie (2004), these two events had the same effect. Because the “competition-for-capital bar” was drifting up, it became necessary to reread the GAAP regulations in order to find creative ways to manage profits – particularly in the short run. (Also see Estes, 1996). People started hearing the term “creative accounting”. However, it became difficult to find enough GAAP “loopholes” to consistently report a relatively competitive bottom line. This forced organizations to move into the shadowy world of profit manipulation, i.e., “cooking the books”, and eventually led to outright manufacturing of the accounting information needed to stay on top of the presumed ever-rising market. The Enron excesses that had created an “Emperor’s new clothes” version of the market were finally disrobbed by the “We have a problem” memo penned by Enron’s VP of Corporate Development, Sherron Watkins, an ex-employee of the ex-flagship of the accounting fraternity Arthur Anderson. At that point, the financial, legal and governmental worlds predictably moved in unison to restore confidence in the markets. This called out specific legislation and regulations, essentially re-writing the conflict-of-interest-rules for audit firms with consulting arms. Further, the resulting legislation required transparent and absolute separation of groups providing IB services from any other related organizations thus preventing the appearance of organizational conflict of interest. A recent example of the result of the regulations umbrellaed under the Sarbanes-Oxley Act (Public Company Accounting Reform and Investor Protection Act of 2002) is the separation of the Salomon Smith Barney “venture”. Another important piece of legislation is from the SEC Reg. 17(B): NASD/NYSE Analyst Regulations requiring that the percentages of Buy, Hold and Sell recommendations made in publicly available IB reports be noted.
The third factor that would eventually affect the BSC was an even simpler, but more deadly, way to service the bottom line that did not require highly paid creative accounting consultation. One simply ignored the municipal, state and federal production regulations, sometimes broadly called the OSHA (Occupational Safety & Health Administration) rules as well as the various jurisdictional pollution and waste disposal guidelines. Organizations such as Hooker Chemical, the Three Mile Island nuclear power facility, Pacific Gas and Electric, Nike, and the Gap, Inc. tried, and for some time, were successful in looking good at the expense of just about everyone. These violations of the public trust, while temporarily providing improvements in the bottom line and resulting in blockbuster movies such as Erin Brockovich and Civil Justice and award winning books such as Ivins (1995) Toxic Sludge is Good for You and Karliner’s The Corporate Planet (1997) were in reality a sad commentary on the results of an obsessive preoccupation with the financial bottom line.

These three effects coalesced in the 1990s. And, as Higgins and Currie (2004) clearly demonstrate, they continue today, pressuring organizations to reduce costs by whatever means necessary, so as to be viewed as a relatively attractive investment alternative in the capital markets. The disconnect between the organization and its societal context caused by obsession with relative profitability also seems to have a positive “behavioural” effect. The disdain that most organizations feel for monitoring and regulation, unarguably a justifiable societal response to the excesses of the 1990s, created an interesting counter-reaction. Organizations began to see the wisdom of taking the actions that are needed to bring their goals and those of their stakeholders into congruence; and thus avoiding the monitoring and oversight that is the predictable result of societal-organizational goal conflict. Enter the BSC, which could be a possible way to deal with the goal-congruence problem often resulting when capital competition is focused solely on profitability.

How and what is the BSC doing currently?

As the BSC is now in its second decade of use, there have been a number of articles, perhaps in response to the dysfunctional consequences resulting from the pressure experienced by organizations, to present a relatively positive market profile, suggesting that the BSC is in need of refocusing. This refocusing seems to have two relatively different dimensions. The first is to again broaden the financial perspective from measures that address internal financial performance to those that are more market-oriented. According to the Hackett Group (2004, p. 67), the majority of Balanced Scorecards are “out of balance because they are overweight with internal financial measures”. Further, according to Robert Paladino (2005), former vice president and global leader of the Telecommunications and Utility Practice for Norton’s company, the Balanced Scorecard Collaborative, http://www.bscol.com/, a major failing of the BSC is the fact that many organizations adopt a piecemeal approach to its implementation that lack the necessary coordination linkages between performance and rewards so as to motivate the intended use of the scorecard information. This opinion also fits well with the findings of Stratton, Lawson and Hatch (2004, p. 39), who note: “... scorecard systems do not universally encompass the use of performance targets, link to reward systems, or provide for feedback to managers. However, when organisations do address these issues, their chances of having a successful experience with score-carding increase.” Paladino recommends that the final evaluation dimension of organizational activities that can help in coordinating all the disparate aspects of organisational performance is to focus on the Maximization of Shareholder Value. This then is the first aspect: a better systematic connection of the organization to the market through the BSC.

The second aspect of the refocusing is again to broaden the scope of the BSC, but this time in the direction of the firm’s stakeholder context by including in the BSC actual non-financial performance measures that address Corporate Social Responsibility (CSR) or Corporate Social Performance (CSP). Such measures are the welfare dimensions; not only the welfare of the shareholders who are concerned with the market performance, but also of the stakeholders of the organization.
In the past, starting only a few years after the introduction of the BSC, researchers began offering their advice that the BSC needs to address the CSR/CSP aspects of organizational activity (Johnson, 1998; Epstein & Wisner, 2001; Swift, Owen & Humphrey, 2001; Dias-Sardinha, Reijnders & Antunes, 2002; Figge, Hahn, Schaltegger & Wagner, 2002; and recently Higgins & Currie, 2004; and Van der Woerd & Van den Brink, 2004). Higgins & Currie (2004, pp. 304, 306) provide the following cogent expression of the necessity to widen the scope in practice of the BSC:

“These scorecards do nothing to translate financial performance into performance of the corporation at a broader level than just the interests of the stockholders and corporate management. Corporate scorecards are badly in need of rebalancing. ... We propose that at a minimum a social responsibility performance perspective becomes part of the business scorecard.”

Van der Woerd & Van den Brink (2004), in a pilot effort, have taken the first steps to integrate measures that go beyond the traditional financial aspects of corporate performance into a scorecard. They have created, as their response to the identified need to rebalance the BSC, what they refer to as the Responsive Business Scorecard (RBS). They note that (Authors abstract citation):

“The European Corporate Sustainability Framework (ECSF) program distinguishes several ambition levels for Corporate Sustainability/Corporate Responsibility. The traditional Balanced Scorecard is suitable for companies that aim for Compliance-driven CS/CR or for Profit-driven CS/CR, where the financial bottom line is the ultimate indicator for success. More ambitious companies want to balance economic, social and ecological targets in a Community-driven CS/CR or Synergy-driven CS/CR. For ambitious companies, we propose a format of a Responsive Business Scorecard (RBS). The Responsive Scorecard enables companies to score at Profit, People and Planet, at the same time to integrate stakeholder demands into internal programs to improve performance.”

It is important to recognise that addressing CSR/CSP issues is not just organizational altruism in action; failure to exercise control and good judgement regarding CSR may have negative consequences for two reasons: (1) CSR/CSP may now be important variables in determining the market value of the firm; perhaps they are market drivers along with the financial drivers of Kaplan and Norton; and (2) recent history teaches that organizational insensitivity to society’s needs usually provokes a broad-based societal reaction resulting in monitoring, regulation, and a burgeoning bureaucracy.

In summary, there have been calls for refocusing and an ambitious pilot effort to begin this process. Refocusing is another way of saying that the BSC has not lived up to its potential. This is exactly what a report issued by the Institute of Management and Administration [IOMA] (2004, p. 4) <www.ioma.com> finds. Consistent with the Hackett report (2004), they report that “70% of all companies that implemented balanced scorecards fail to generate real business value through their use.” Consider now the implications of integrating CSR/CSP measures into a scorecard.

The BSC and CSR/CSP measures

The apparent dearth of utilization of CSR/CSP measures in the BSC may speak to uncertainty as to how to measure these dimensions. One may, justifiably, cite such measurement difficulties as a reason for organizations failing to embrace stakeholder issues during the 1970s (See Davis, 1973). However, currently such measurement difficulties are no longer an acceptable rationalization because in 1988 Kinder, Lydenberg, Domini & Co. (KLD) of Boston offered a CSR/CSP measurement system and their independent assessment of corporate activity. We will use the KLD measures in our study for the following five reasons: (1) They were the first group to establish a comprehensive, fully articulated, and transparent measurement system. (2) They make summaries of their evaluations publicly available. (3) Their detailed evaluations are a part of major data sources such as the Wharton Research Data Service, and so detailed CSP information is also available. (4) They were one of the first to establish a social responsibility (SR) index, the Domini 400 (DS 400). (5) They
have recently partnered with NASDAQ to form an annually re-constituted daily priced SR index using 350 NASDAQ stocks called the KLD NASDAQ® Social Index (KLD NS).

Therefore, given their efforts over almost two decades, it now is possible to measure effects of the execution of corporate policy that extend beyond the usual financial and “quasi-financial” dimensions of the BSC.

However, there is one further critical aspect that needs to be considered in order to rationalize CSR/CSP measures as a BSC enrichment, i.e. the likelihood that they will be integrated into the BSC. It is unlikely that the BSC will be utilized to plan, monitor and evaluate the social responsibility effects of the execution of organizational policy if the market reacts adversely to organizational attempts to address CSR issues. We have learned, the hard way, the persuasive, often compelling, and sometimes corrupting power of the possibility of economic gain. Consider now the reported research results of the market effects of organizational policy actions as viewed through the KLD measures.

**KLD and the Market Orientation**

Here, we are considering only the market effects of organizational CSR/CSP because market value is not a derivative measure and it is a direct link to shareholder wealth. As Hillman & Keim (2001, p. 133) note, “Conceptually, MVA is the closest operationalization available to us to capture our dependent variable of interest: shareholder wealth creation”. One supposes that other measures such as GAAP-measured profitability, residual income and its re-labelling as EVA™ and cash-flow affect the market and so may be related, in varying degrees, to market value. However, the “final-bell” market value is the only objective measure of the change in shareholder wealth.

**Research Reports on the KLD/Market connection**

To evaluate the relationship between the market and the KLD measures, we conducted the following search using the ABI/Inform™ and Business Source Premier™ databases on 15 May 2005: KLD AND [beta OR Jensen OR alpha OR Sharpe OR SPI OR Treynor OR TPI OR CRSP]. In addition, we searched the ISI Social Sciences Citation Index™ and Econlit™ databases using just the initials: KLD. This resulted in 40 articles, eight of which presented market-related studies. The first study identified by our search using KLD measures relative to the market effects was done by Johnson & Greening (1994). They investigated the way that the KLD measures of Community, Minorities, Environment and Product as scored by their factor structure – a correlation-combination of the various issue-screens – related to a variety of corporate-profiling measures drawn from 1990 proxy statements. In addition, they used the following market measures: Jensen’s $\alpha$, the Sharpe Performance Index [SPI], and the Treynor Performance Index [TPI]. We will discuss these measures in some detail following. Johnson & Greening (1994, p. 318) found, respecting market value associations, that “the predicted path between market performance and the community/minority dimensions of CSP was not supported”. Essentially, for their corporate social performance measure they did not detect a market effect, although they do note that there was an effect for GAAP accounting performance. We wanted to briefly review this study because we will be referring to it as a basis of comparison for our results.

The results of the remaining seven studies may be efficiently summarized by considering them in two groups. (1) Results on SR indices/portfolios, where one examines the temporal market performance of these indices/portfolios, such as the DS 400, compared to the market on a variety of measures, usually $\beta$, the SPI or the TPI. (2) Results on firm studies, where the market returns of a number of firms classified as socially responsible on the KLD or derivative measures are compared to firms classified as not socially responsible. The definitions of the CSR/CSP classifications vary considerably over these studies. We view this lack of definitional uniformity positively, in that it gives a certain “robustness” to the reported results.

1. The SR Indices and Portfolios: Here one does not need a meta-analysis to draw an inference. We found five studies by Hopkins (1992), Kurtz (1997), Johnson & Greening (1999), Guerard & Stone (2002), and Boutin-Dufresne & Savaria (2004); all of which essentially find the same effect over the various definitional frames. We have selected Boutin-Dufresne & Savaria’s (2004, p. 64) summary to characterise these study results:
There is growing evidence suggesting that market participants care – or at least should care – about corporate social responsibility. The literature has mainly examined the issue of a firm’s social behaviour in relation to its financial performance. In this article, we look at the question in relation to financial risk instead. More specifically, we looked for a relationship between idiosyncratic risk and the level of social responsibility derived from a rich database on the social performance of Canadian firms. Results according to two different methodologies support social investors’ views on SRI. That is, investing ethically does not seem to impair the financial prospects of a portfolio. Our findings should be of interest to both CEOs and investors. For the former, our results indicate that the adoption of socially responsible codes of conduct could help reduce the overall business risk, as defined in financial theory, of their corporation. For the latter, such an initiative could also improve the risk-adjusted stock performance of ethical firm portfolios in the long run. From a portfolio management standpoint, our findings indicate that combining socially responsible stocks into portfolios could reduce their diversifiable risk component. Our results add to the growing evidence that supports the view that socially responsible investing does not hurt portfolio performance.”

2. Firm Studies: The two firm studies, in addition to the Johnson & Greening study discussed above, are:

Hillman & Keim (2001, p. 133): “Our results using MVA as a measure of shareholder wealth creation indicate a positive relationship with stakeholder management and a negative relationship with social issue participation. Our results also indicate that the direction of causality is from stakeholder management/social issue participation to shareholder wealth creation/destruction. Additional analyses support this directional causality in that the reverse causality is not statistically supported. Thus, our findings are consistent with our theoretically based predictions that stakeholder management can lead to shareholder wealth creation and that participation in social issues does not lead to shareholder wealth creation”.

Mattingly (2004), like Johnson & Greening, did not find a market effect.

Therefore the results are mixed; Hillman finds that saliency may make a difference while the other two studies find no effect.

Although the evidence is rather sparsely distributed over the last ten years, one can observe a generally positive result. Social responsibility as expressed through the KLD measures seems to be either associated with a positive market effect or not associated with a negative effect. Consider now our study that also addresses the relationship between market effect and the CSR/CSP as expressed through the KLD issue-screens.

Research Design

Introduction and Statement of Hypothesis

We hypothesize that the market will respond positively to organisations that have exhibited socially responsible behaviour. Specifically,

Hypothesis: Organizations that are rated as socially responsible will have superior market performance compared to organizations rated as not socially responsible.

Information from the eight studies reviewed above, in addition to the following four factors, was used to formulate our hypothesis.

1. Increased awareness of the importance of CSR in the global arena. Simms (2002, p. 49) notes, in an excellent review article, that according to a 2002 survey conducted by PricewaterhouseCoopers that “70% of global chief executives believe CSR is vital to their companies' profitability, while KPMG's International Survey of Corporate Sustainability Reporting 2002 found that 45% of the world's largest 250 companies now produce environmental and social reports, up from 35% in 1999”.

2. Self Monitoring. There are a number of organizations, largely in the financial accounting and consulting SIC codes, that are marketing their expertise as to how to be transparent. For example, consider PricewaterhouseCoopers, which offers this service as ValueReporting™.
They note on their Website: <www.pwc.com>:

“The current financial reporting model is struggling to meet some of its most basic objectives. Simply put, it no longer provides the information necessary to evaluate the quality or sustainability of corporate performance, to differentiate good management from bad, luck from skill. For almost a decade, PricewaterhouseCoopers has invested in a research initiative, known as ValueReporting, which has focused considerable time and effort in understanding what information is important for communicating the performance of companies across the whole spectrum of assets that underpin value and growth today”.

Another organization of note providing such services is: AccountAbility. They offer social reporting training and assistance, and their impressive list of clients includes many of the major international accounting firms such as Deloitte, Ernst and Young, KPMG, and also Nike, the Body Shop, and the Gap, Inc., the last three of which have run into social responsibility difficulties in the recent past. This fits well with the report from Clikeman (2004, p. 27), that 600 firms have issued sustainability reports in 2002; for example, Conoco, Dow Chemical, Procter & Gamble, Siemens and Shell Petroleum, to note a few.

3. Agencies reporting firm CSR/CSP information. Here there also has been a dramatic increase in the number of agencies, groups and organizations that evaluate organisational performance and make their assessments publicly available. Prominent among them are: KLD Research and Associates, Global Reporting Initiative, Ethics Resource Center: Corporate Reputation Watch, Best Practices LLC, Human Rights Campaign, American Customer Satisfaction Index, Fortune100 Best Businesses to Work For, and Business Ethics’ 100 Best Corporate Citizens. And, according to Barron’s, (14 March 2005, p. 11), there is now a PBS television show called the “Ethical Market Place” that will air thematic programs on good corporate citizens. These organizations serve a critical role of providing independent assessment information.

4. CSR/CSP Trading Indices. Finally, there are now a number of SR thematic indices. One of the first such indices, launched in May 1990, was the KLD Domini 400 Social Index (DS 400) that was developed by Domini, one of the co-founders of the KLD group. Others of note are Pax World Balanced (PAXWX), Dow Jones Sustainability Index (DJSI), Pan European Sustainability Benchmark (DJSI STOXX), WilderHill Clean Energy Index (ECO), FTSE4Good Indices, Domini Social Equity (DSEFX), and the recently launched KLD NASDAQ® Social Index (KLD NS). Not all of these indices are exhibiting strong performance. For example, the Domini Social Equity (DSEFX), a one-plus-billion-dollar fund, has slipped badly of late, falling more than 40 places in its Morningstar fund ranking, compared to its 10-year average of 40th place. However, the DS 400 continues to be a strong performer, outperforming the S&P for 16 consecutive years.

There is another side to social performance. There is also a “vice” fund (VICEX), launched in August 2002 and consisting of organizations involved in the alcohol, defence, gambling and tobacco industries. The VICEX, a relatively small fund, has enjoyed noteworthy success of late. According to Reuters (Boston), Vice Sells: (30 April 2005),

“And with tobacco, alcohol and gambling stocks on a roll, Mutuals.com’s Vice Fund (VICEX) has jumped into the top 1 percent of more than 700 mutual funds in its classification. Since the end of last year, the fund has attracted more than $12 million in new money, a 55 percent increase. It now has $34 million under management, and expects to hit $100 million by the end of 2005.”

These vice industries are in reality “Janusesque” in nature. For example, consider defence: the USA now has a Department of Homeland Security and defence, in the USA, has a very different meaning since the WTC event. Also, gambling has been legalized in almost all industrialized countries; such lotteries, and intra-state gambling/casino-zones in the USA provide vast revenues. A significant percentage of those revenues are redistributed to programs such as education, elder-care and general health assistance for low-income individuals. Finally, alcohol has been a beverage of choice for millennia and there is credible scientific evidence that, in moderation, consumption of beverages such as wine and beer do have medicinal benefits. For this reason – i.e., the inherent difficulty in deciding if Defence, Gambling and Alcohol contribute, on net, positively or negatively to the quality of life – we will exclude them in the definition of CSP.
In summary, we offer the above in support of the hypothesis that there will be a positive market effect relative to CSR/CSP behaviours as expressed through the KLD measures. Consider now the research variable groupings.

**The CSR/CSP Research Variables**

KLD offers a rich menu of stakeholder measures that one may use to extend the scope of the BSC. KLD evaluates the performance of organizations on the following 16 general dimensions called issue-screens: Community, Corporate Governance, Diversity, Employee Relations, Environment, Human Rights, Product, Abortion, Adult Entertainment, Alcohol, Contraceptives, Firearms, Gambling, Military Weapons, Nuclear Power and Tobacco. The first seven of these dimensions are each subdivided into **Strengths** and **Concerns**. For example, according to the KLD Ratings Data Inclusive Social Rating Criteria (2003, p. 3) report for the Community issue-screen, the **Strengths** or positive aspects include:


The Concerns, or negative aspects, include: Investment Controversies, Negative Economic Impact, Indigenous Peoples Relations and Other Concerns.

Each of these Strengths or Concerns has specific definitions. For example, Generous Giving is defined as follows:

“*The company has consistently given over 1.5% of trailing three-year net earnings before taxes (NEBT) to charity, or has otherwise been notably generous in its giving.*”

The remaining nine screens are somewhat difficult to place as **Strengths** or **Concerns**. For example, as discussed above, Alcohol, Defence [Firearms and Military Weapons], Gambling, as well as Nuclear Power or Contraception, all have both positive and negative aspects, depending on one’s point of view. Therefore, we will concentrate in our study on the first seven measures:

Community, Corporate Governance, Diversity, Employee Relations, Environment, Human Rights, Product

Also, it is important to note that these measures are not specifically identifiable or categorized as to their saliency (Agle, Mitchell & Sommerfeld, 1999). Therefore, depending upon the specific organization, any one of these issue-screens may have identifiable stakeholder interests or may be more of a “free-floating” societal issue. While we believe that saliency can be important, particularly insofar as governance is concerned, creating an enterprise stakeholder saliency map is a formidable measurement challenge that merits research attention. Finally, we are, of course, not implying that all organizations evaluated by KLD use the BSC only that organizations that are sensitive to truly non-financial measures in the CSR/CSP context must collect such evaluation information and then allocate the resources necessary to address the issues to which the organization is sensitive. This is essentially how the four measures that currently constitute the BSC are supposed to be used.

**The Market Variables and Data Sources**

To examine the relationship between market performance and the seven above noted KLD issue-screens that we propose for the BSC, we collected daily return information from the Center for Research in Security Prices (CRSP®) service for the 647 publicly traded organizations that were part of the KLD report *Corporate Social Ratings for 2002* published 18 March 2003. To focus the study on the market effect of the KLD social performance audit, we collected daily market returns from the date of publication of the above report until the end of 2003. To examine the various ways in which the organizations’ KLD social performance ratings impact the market, we will use the following market measures:

1. **The CAPM Measures.** The first measure, also used by Johnson & Greening (1994), is Jensen’s $\alpha$. Jensen’s $\alpha$ (sometimes just called $\alpha$) is the difference between the average rate of return of a security or portfolio and its SML; i.e., the CAPM risk-return line. Computationally, it is the intercept of the excess returns regression, and thus a measure of excess performance relative to the risk-free rate. A positive (negative) Jensen’s $\alpha$ indicates that the company outperformed (was outperformed by) a random market portfolio. The CAPM Beta ($\beta$) is a measure of relative varia-
tion, i.e., co-variation of the company’s return with the market’s returns to the variation of the returns of the market. Therefore, β is a risk-return measure of the organization relative to the market, assuming that a variation-based measure of return surrogates for risk. If β is greater (less) than 1, the company has more (less) risk relative to the market, i.e., higher (lower) relative-return variation (Bodie & Merton, 2000). These measures are un-indexed relative-to-the-market performance measures for the firm. As is common practice, we are using the S&P500 value-weighted index as the market surrogate. For the risk-free rate, we are using the 30-day T-Bill composite.

2. The Sharpe and Treynor Performance Indices. These indices, also used by Johnson and Greening (1994), are risk-indexed excess return measures. The Sharpe Performance Index (SPI) is the excess return of the organization relative to total risk as measured/surrogated by the standard deviation of the returns of the organization. Computationally, the SPI is the average return of the organization less the average risk-free rate, here the T-Bill composite, divided by the standard deviation of the organization’s returns for the time period in question. The Treynor Performance Index (TPI) uses the same numerator as does the SPI, but divides it by the firm’s period β. In this sense, the Treynor index measures excess return as the ratio of excess return to the non-diversifiable or systematic risk as indexed by β. Thus the SPI and the TPI present risk-indexed excess return information.

3. CRSP® Standard Deviation (σ) and Beta (β) peer groups. The excess return of an organization may also be measured relative to the average return of a peer group. The CRSP® services report such excess return information for the following two peer groups: a Standard Deviation (σ) or total-risk peer comparison group and a Beta (β) or systematic-risk peer comparison group. Computationally, these measures subtract from the organization’s daily return, the average return of the organizations that are in its daily peer comparison group (CRSP®, 2005, Definition Macro). For example, consider the β-peer group. The CRSP® service groups all of the organizations for which it collects data into a number of clusters based upon the range of β. Then, for each cluster, the mean return is computed and subtracted from the return of each organization in that β-peer cluster. In this sense, organizations with a positive (negative) average for their β-peer group have on average outperformed (been outperformed by) their β-peers. We have included these measures because they provide a strong test of relative excess-return performance, in that they are measured against excess return of a peer group that over time must exceed the risk-free rate. Given that we now have identified the variable sets both for the CSR/CSP and the market dimensions, consider the specific measurement protocols for the KLD issue-screens.

The organization of the KLD information

There are myriad numbers of ways to organize the 647 study organizations into CSR/CSP groups. We discussed the various possibilities with Noel Friedman, CFA, Director of Product Management & Development at KLD Research & Analytics, Inc. After these fruitful discussions, and considering the Johnson & Greening studies cited above, we arrived at the following way to score and group the study organizations and organize the analysis.

We will use the difference between the number of Strengths and Concerns as the CSR performance measure of an organization individually for the following seven issue-screens:

- Community
- Corporate Governance
- Diversity
- Employee Relations
- Environment
- Human Rights
- Product

For example, in the KLD 2003 report mentioned above the IBM Corporation was scored by KLD for the Diversity issue-screen as having six Strengths and no Concerns, and for the Employee issue-screen IBM had one Strength and two Concerns. Therefore, for the Diversity issue-screen IBM would have a KLD difference-score of (+6) and for the Employee issue-screen they would have a (-1) as the difference-score. We did not weight particular Strengths or Concerns as to relative importance.

Because the number of Strengths and Concerns measured for each of the issue-screens differs and the symmetry of the distribution of the scores also varies over the issue-screens, KLD urges caution in drawing inferences by using an aggregate KLD score, i.e., a score calculated over all of the criteria. Therefore, to avoid asymmetrical-weighting and averaging problems, we created two CSP groups for which we will measure market performance. The first grouping is those organizations that had no negative KLD difference-scores on any of the seven issue-screens, and at
least one positive KLD difference-score recorded among the seven issue-screens. We will note members of this group as the pure-relative-strength organizations group (PRSO, n = 87). At the other end of the spectrum, we created a group for which there were no positive KLD difference-scores recorded, and at least one negative KLD difference-score among the seven issue-screens. We note this grouping as the pure-relative-concerns organizations group (PRCO, n = 116). Using this partitioning a few organizations, 12 out of the 647, were eliminated from the analysis because they did not have at least one negative or positive KLD difference-score. That is, they were either not rated on some of the seven screens or they had an equal number of Strengths and Concerns on particular issue-screens, so that they had zeros as their difference for all the seven issue-screens. The other group that this PRSO and PRCO partition creates is a mixed group of 432 organizations for which there was a mix of positive and negative KLD difference-scores. We will contrast only the pure groups in examining the CSP effect because if we expect to see a market effect, this partition is where it should be observed. We will briefly present the orientation of the mixed organizations relative to the PRSO and PRCO groupings.

In summary, we will test for market performance differences between the PRSO and PRCO groupings for each of the six market measures: $\beta$, Jensen’s $\alpha$, SPI, TPI, $\sigma$ Peer Group and $\beta$ Peer Group. Conservatively, for these statistical tests, we will report the largest two-tailed p-value as between the appropriate parametric and non-parametric tests. Finally, considering the hypothesis, specifically, we expect that the PRSO will outperform the PRCO on the six market measures.

Results

Consider first the factor results of the study. We are interested in comparing our results to the factor structure developed by Johnson & Greening (1994). Recall that they used KLD difference-measures on Community, Minorities (this issue-screen has been currently redefined from the 1990 version used in their study to Diversity), Environment and Product. To better understand the factor results of our study, we added the Employees Relations issue-screen.

The Factor Study Results

To study the factors here, we are using the organizations that had KLD ratings for all five of the above-noted issue-screens, n = 47. The factors were produced by using a principle-component-varimax rotation on the five variable-issue-screen associations, as measured by the Pearson product-moment correlations for the unweighted KLD difference-scores. Because the third ordered eigenvalue was less than 1.0, we used a two-factor rotation. This produced, as the first factor, the Community and Diversity issue-screens that were positively associated, each had a factor-variable loading greater than $\sqrt{0.5}$. For the second factor, the Environment and Product issue-screens also were positively associated and had factor loadings of 0.712 and 0.694 respectively. The Employee Relations screen was distributed essentially equally between the two factors; it was positively associated with the first factor and negatively associated with the second one (Fig. 1).

Fig. 1. Factor Relationships for the five KLD issue screens

This factor structure is basically identical to that reported by Johnson & Greening (1994, p. 316) who report that “two factors may be present within the corporate social performance construct namely, a community/minorities and a product/environment factor.” Because our two-factor structure shows remarkably good loading-separation and stability with the Johnson & Greening
results, it seems warranted to discuss briefly the associational relationships sketched out by our two-factor-profile.

1. Community/Diversity/Employee Relations Factor. This factor seems to be a “localised-societal” (i.e., community) dimension where the Community, Diversity and Employees relative Strengths and Concerns move in concert. That is, Strengths > Concerns group together, as do the negative scores (Concerns > Strengths), suggesting that organizations that have taken actions to be sensitive to Community and Diversity issues also take care to treat their employees well and vice versa.

2. Environment/Product/Employee Relations Factor. The second factor seems to be the environmental/market context for the organizations. The unexpected negative association of the Employees issue screen in this factor merits further investigation. This negative association indicates that organizations that are trying to “do the right thing” concerning the environment and their product have a negative profile regarding their employees. And, at the other end of the associational split, organizations with more Concerns than Strengths regarding the environment and product are those that have been rated for the Employee Relations issue-screen as having more Strengths than Concerns.

As a conjecture to provide an interpretation of the former relationship, we offer that where market value seems to be positively affected both by announced: downsizing, outsourcing, or “Saturday Night Massacres”; as well as by actions taken to address CSR/CSP issues, that, ironically, in pursuit of superior market performance organizations may be attending to the CSR/CSP dimensions while willingly sacrificing their employee base. An example of this “Jekyll and Hyde” confliction is Deutsche Bank. On the one hand, they seem to have a good CSR/CSP profile. According to Cheryl Chickowski (2004, p. 1), Head of Sponsorship and Hospitality at Deutsche Bank AG, “Deutsche Bank spends about 50% of its sponsorship budget on the arts”. For example, in 2005 Deutsche Bank AG, which holds the world’s largest corporate collection of fine-arts, presented a 25-year-jubilee-exposition of art selected from this collection at the Deutsche Guggenheim Museum in Berlin, Germany. Further, Deutsche Bank was listed on the Pan European Sustainability Benchmark (DJSI STOXX, May, 2005). Finally, according to Sargent (2005, p. 1) Deutsche Bank, AG gave 10 million € for the tsunami-relief-effort which was “more than four times the commitment made by many of its European and American counterparts”. On the other hand, after having announced record-breaking profits for 2004, in a February 2005 press conference Deutsche Bank president, Josef Ackerman, announced that there would be a layoff of 1,920 employees in Germany. This downsizing announcement was met with a market bounce (Reuters News, 11 February, 2005).

Needless to say, there can be many underlying scenarios consistent with such associational relationships. However, for our purposes, we wish only to note that the emergence of a strong, temporally consistent factor structure is an important piece of research information. It suggests, in a statistically convincing way, that finding by chance a factor structure with such temporal-similarity is most unlikely, and therefore one may assume that there is some underlying structure tapped into by the KLD difference-score data. This is another way of saying that if no discernable associations could be found among the difference-scores for these variables, so that a factor structure did not emerge, this might call into question the interpretation of the various variables over the market measures.

The Market Results

Consider now the results of the six market measures as viewed through the market performance of the PRSO and PRCO, as presented in Table 1.
For Table 1, due to the fact that the scales of the variables are different, we have presented the results as ratios of the PRSO to the PRCO. For example, considering $\beta$, the CAPM risk-relative measure of the firm, the mean $\beta$ for the 87 organizations in the PRSO group was 0.740385; and for the 116 organizations in the PRCO category, the mean $\beta$ was 0.770620. This gives a ratio of 0.96, and the p-value of this difference in means was 0.45, suggesting that the difference in means between the two groups was not sufficiently large, to reject the null that there is no difference in $\beta$ between the two groups. This result fails to support the hypothesis that the PRSO will outperform – i.e., have a lower $\beta$ than the PRCO. However, this result is consistent with those of Boutin-Dufresne & Savaria and Kurtz, who found that socially responsible firms did not pay for their socially responsible decisions in terms of their relative risk-return profile – i.e., have a higher $\beta$.

Consider now the interpretation of the results presented in Table 1 for the remaining five market measures all of which support the hypothesis.

1. **Jensen’s $\nu$.** The ratio of the means of Jensen’s $\nu$ of the PRSO to the PRCO is 2.91 which has a p-value less than 0.01. This suggests that the organizations that have taken actions to be sensitive to the issues concerning Community, Corporate Governance, Diversity, Employee Relations, Environment, Human Rights or Product were on average rewarded with a higher excess return as measured by Jensen’s $\nu$ – i.e., relatively outperforming the market-portfolio.

2. **The SPI and TPI.** The results for these indexed return measures are essentially the same. On average, the PRSO outperformed the PRCO; both with respect to excess return relative to total risk, i.e., the SPI, as well as relative to systematic risk as indexed by $\beta$ i.e., the TPI. One should note here that the result for the TPI is a pure-excess return difference because, as noted above, there was no statistically significant difference for the mean of the $\beta$s for two groups. This is important because if there was a significant difference in $\beta$, then that would affect the TPI relative comparison, that is a relative-$\beta$ measure.

3. **CRSP® $\sigma$ and $\beta$ peer Groups.** These peer-comparison results suggest that the PRSO are outperforming, on average, the PRCO for both of their peer-comparison groups. This is consistent with the SPI and TPI results.

We conducted the same analysis for the mixed-group of organizations – i.e., those organizations that have both Strengths and Concerns in their profiles. For this group of 432 organizations, we found, not surprisingly, according to the hypothesis justification discussed above, that their scores are in between the scores of the PRSO and PRCO for all of the market measures that support the hypothesis tested. As further research, one could investigate the trade-off sensitivity between the KLD Strengths and Concerns by issue-screen relative to these market measures.

**Summary and Conclusion**

**Summary**

The following simple market-pattern emerges that is generally consistent with the hypothesis: Organizations with a strong KLD profile seem for the most part to be reaping market rewards, compared to those organizations which have policies that have raised concerns as ex-
pressed through the following seven issue-screens: Community, Corporate Governance, Diversity, Employee Relations, Environment, Human Rights, or Product. This differential reward appears in all five of the relative-profitability market measures used in the study, and to this extent suggests that these CSR/CSP KLD issue-screens are also market drivers. Thus, they are logical candidates for inclusion in the BSC.

Conclusion

In response to the calls for refocusing and rebalancing the BSC, and based upon the above reported study results, we recommend the following BSC enrichments:

I. To address expanding the BSC to be more sensitive to shareholder wealth creation, we suggest adding the following three classes of market measures to the BSC:
   1. CAPM Measures: Jensen’s $\alpha$ and Beta ($\beta$).
   2. The Sharpe and Treynor Performance Indices.
   3. The CRSP® $\sigma$ and $\beta$ Peer Comparison Groups.

II. To address expanding the focus of the BSC to be more sensitive to the CSR/CSP profile of the organization, we suggest including appropriate KLD issue-screens in the BSC.

We have come full circle. The “non-financial” measures first proposed by Kaplan and Norton: Customers, Internal Business Processes, and Learning and Growth really focused on improving the financial bottom line and so addressed only shareholder wealth. In use, however, the BSC was found to be heavily weighted in terms of the internal financial dimension, and also lacking a stakeholder context, and so it was criticised as being out of balance and in need of refocusing. Ironically, adding truly CSR non-financial measures of performance may provide a synchronous connection to the market because these non-financial measures turn out, in market terms, to be important financial signals to those who are driving the market. That is, the true market drivers are the non-financial measures.

Having begun the paper on a practical note, we wish to end it on a theoretical one. We see here a realisation of the various aspects of Milton Friedman’s succinct statement of the role of the organization as published in the popular press: The New York Times Magazine (1970, p. 17). He stated that the organization’s resource conversion process must be aligned with “shareholders’ desires, which generally will be to make as much money as possible while conforming to the basic rules of society, both those embodied in law and those embodied in ethical custom.” To restate this in a more contemporary context, being sensitive to the institutional framework (Williamson, 1996) within which organizations operate is “good business.”

Integration of transparent stakeholder measures into the BSC makes theoretical sense because it is connected in a clear way to the theory of agency and its related moral-hazard issues, given the possibility of information asymmetries. As Aronsson & Lofgren (1996), Aronsson & Wikstrom (2003), Andersson, Aronsson & Wikstrom (2004), and Aronsson (2005) clearly demonstrate, when there is publicly available transparent information that is known to all parties in the principle-agent setting; and the cost of expected sanctions is greater than the possible defection gains, then principle-agent goal congruence is to be expected in the long-run where there is exit and entry fluidity in the factors-of-production market. Here the disdain for monitoring and regulation seems to provide the sanction expectation. There is no doubt as to what happens when societal-well-being is threatened as the avalanche of regulations caused by Enron, Tyco, Quest, Computer Associates, Global Crossing and WorldCom clearly demonstrates. The goal congruence, as guided by the market, is of course contingent on the relevance and believability of the measured stakeholder information (Herbohn, 2005 p. 534). The presumed pressure exerted by the possibility of independent regulation oversight given the provision of independent and transparent CSR/CSP information such as that provided by KLD seems to provide goal-alignment between the principal, here society – i.e., us and the agents, here the organizations – i.e., us again playing a different role.
References


113