

Relations among Measures, Climate of Control and Performance Measurement Models

Mary A. Malina

University of Vermont
mmalina@bsad.uvm.edu

Hanne S. O. Nørreklit

Århus School of Business
hann@asb.dk

Frank H. Selto

University of Colorado at Boulder
frank.selto@colorado.edu

October 5, 2006

Forthcoming in *Contemporary Accounting Research*

We appreciate comments and suggestions from workshop participants at Rice University, University of Vermont, San Diego State University, Århus School of Business, University of Tilburg, University of Colorado-Boulder, University of Ghent, the Management Accounting Research and Case Conference 2005 and the North American Field Research Conference at Queen's University 2005. We thank Qihong Zhao, Yanhua Yang and Veronda Willis for research assistance. We gratefully acknowledge significant contributions from an associate editor and two anonymous reviewers.

Relations among Measures, Climate of Control and Performance Measurement Models

ABSTRACT

Cause-and-effect relations among performance measures are alleged to be distinguishing features of performance measurement models (PMM), such as balanced scorecards. This study reports the evolution of the study of a PMM that was developed by a large U.S.-based company for its closely linked distribution channel. Motivated by the literature on PMM and causality, we report an analysis of linked performance measures for 31 quarters (1997 – 2005) and 31 business units. We find minimal statistical significance and no significant predictive ability in the model (i.e., no Granger causality), yet the company and its distributors express satisfaction with the model and with both company and distributor profitability. Reasoning that cause and effect was not the only explanation for scorecard success, we thoroughly analyze qualitative data for how managers perceived and used (a) the relations in the scorecard and (b) the climate of control intended and achieved in the organization through the scorecard. We find that the PMM's logical and finality relations support the company's climate of control. We also find qualitative evidence that the use of the PMM creates an effective climate of control. We tentatively conclude that effective management control does not require statistically significant cause-and-effect relations in a PMM when other factors create a strong climate of control.

Key Words: performance measurement model, balanced scorecard, cause and effect, management control

Relations among Measures, Climate of Control and Performance Measurement Models

INTRODUCTION

This study reports the evolution of an investigation of the cause-and-effect properties of a performance measurement model (PMM) developed by a *Fortune 500* company for its North American distribution channel. The company and this study refer to the PMM as the Distributor Balanced Scorecard or DBSC. The study follows previous, related research and also is motivated by balanced scorecard literature that stresses the importance of the cause-and-effect properties of balanced scorecards (e.g., Kaplan and Norton, 1996, 2001; Ittner and Larcker, 2003) and empirical research that suggests causality (Bryant *et al.*, 2004; Ittner *et al.*, 2003; Banker *et al.*, 2000; Ittner and Larcker, 1998; Rucci *et al.*, 1998). Previous research by Malina and Selto (2001, 2004) has established that the company implemented the DBSC to communicate and match its new customer-service strategy, to provide a more diverse, accurate and balanced set of performance measures, and to direct distributors' decision making. Extant research implies that a well specified PMM reflects a firm's production function and that cause-and-effect relations among measures drive control effectiveness. We review how cause-and-effect relations among performance measures are beneficial for control purposes. The present study then proceeds as an econometric validation of the cause-and-effect properties of relations among measures of the DBSC. However, refutation of cause-and-effect in the DBSC lead to consideration of alternative explanations for the company's continued use and professed satisfaction with the DBSC. These plausible, internally consistent alternatives provide motivation for future research that might support cause-and-effect properties in other PMM, the alternative explanations, or both.

Research Questions

This study begins with the following research question:

A. *Do DBSC relations from the distribution strategy map exhibit valid cause-and-effect properties?*

We analyze the company's distribution strategy by investigating company documents and transcripts from interviews with five distribution managers and DBSC designers and nine distributors. We initially look to these data for evidence of causality in its DBSC. From the qualitative data, we document perceived linkages among the DBSC's performance measures that were validated by managers. The elicited strategy map generates testable, cause-and-effect relations, which are described in detail later. We test these relations using 31 quarters of performance data (1997 – 2005) and multiple tests for cause and effect. Overall, few hypothesized leading-performance measures in the DBSC explain lagging measures, and none of the estimated model relations containing hypothesized performance drivers has significantly better predictive ability compared to models containing only lagged dependent variables (i.e., causality

tests by Granger, 1969, 1980). Yet despite the refutation of causality by empirical tests, the company and its distributors express satisfaction with the DBSC and plan to deploy it worldwide. Hence, we continue the study with a second research question:

B. Are statistically significant cause-and-effect relations necessary for effective management control?

We consider alternative explanations for the apparent ongoing success of the DBSC through the lens of management control theory. To do so, we expand our qualitative data through additional analyses, interviews, and review of company documents (i.e., data not in Malina and Selto, 2001). Importantly, additional qualitative analyses revise our prior conclusion (Malina and Selto, 2001), and we find that the relations among performance measures perceived by DBSC users are *not* cause-and-effect relations. In addition to the previously supported communication benefits, the re-analyzed qualitative data provide evidence that managers and distributors regard the DBSC as an effective management control because its communicated relations among measures create a complementary (1) credible story of success, (2) reinforcement of the company's pay-for-performance culture, and (3) result control that is legitimate and fair. The company has used the results of the DBSC to guide consolidation of distributorships from 31 to 19, and continuing managers appear to alter strategic and operational choices consistent with the DBSC measures, both without statistical evidence of reliable cause-and-effect relations among the measures.

We conclude that managers' beliefs about relations support the organization's climate of control and drive the design and continued use of the DBSC. We also tentatively conclude that statistically valid cause-and-effect relations may be unnecessary to achieve desired control effectiveness in this context and perhaps in others. While this result seems surprising in light of the normative PMM literature, the expectation of cause-and-effect relations may reflect common assumptions rather than evidence. Organizations may use dynamic PMM that are composed of relations that are not cause and effect, but may be more than common sense, to facilitate strategic communication and to create a climate of control rather than to create a predictive business model for use as a decision aid, business simulation, or input-output model (e.g., Zimmerman, 1997: 4-5). Perhaps a predictive business model is the least important reason for a PMM.

This study next reviews relevant cause-and-effect relations literature. The study then reports the qualitative modeling of cause and effect in the DBSC and, next, econometric efforts to refute cause and effect, which were successful. The study proceeds with the evolution of the inquiry by developing plausible alternative explanations. The econometric results and alternative explanations challenge common assumptions about the existence and importance of cause-and-effect relations in PMM. These tentative conclusions can serve as points of departure for future research.

IMPORTANCE OF CAUSE AND EFFECT IN PMM

Proponents of PMM invariably cite their inherent cause-and-effect relations as a major source of the value of such models. We wish to precisely define what we mean by cause and effect because it is not clear that all PMM researchers use a common definition. Most scientists and theories of science adopt Hume's criteria for a cause-and-effect relation (Cook and Campbell, 1979; Edwards, 1972, vol. 2: 63; Slife and Williams, 1995; H. Nørreklit, 2000), and this study also adopts them. The criteria, which are restrictive, are (1) independence, (2) time precedence, and (3) predictive ability. The independence criterion states that events X (the cause) and Y (the effect) are logically independent. Further, one cannot logically infer Y from X but only can do so empirically. The time-precedence criterion states that X precedes Y in time, and the two events can be observed close to each other in time and space. The predictability criterion is that observation of an event X necessarily implies the subsequent observation of the other event Y.

Cause-and-effect relationships are well known in physical sciences and likely exist in firms' physical production functions. For example, a cause-and-effect relationship exists between applied heat and the temperature of water. The heat of a fire and the temperature of water are independent phenomena, and a rise in water temperature occurs after the application of heat. Furthermore, one can predict the water's future temperature from the observed rate of heat transfer using a theoretically based, cause-and-effect relationship. Similarly, firms in many industries may develop PMM that (partly) reflect underlying physical processes.

Benefits of Cause and Effect in PMM

For several decades the strategic management literature has presumed the existence of cause-and-effect relations among key performance indicators (KPI) or measures at various levels of the firm.¹ Although physical processes such as those in chemical industry are analogous to heating water, many KPI relations can be more complex and less deterministic. Nonetheless, the notion of cause and effect among KPI is widespread. For example, Porter (1985) revolutionized strategic management with the application of the value chain concept, which links KPI along the product and service delivery chain. Kaplan and Norton (1992) introduced the notion of a causal balanced scorecard, which has influenced the management accounting literature and which is a direct descendant of the value chain and systems models.² These seminal works argue that cause-and-effect relations exist among proper KPI, and all of the supporting literature identifies process and outcome benefits from building PMM with cause-and-effect relations. We

¹ Frigo, (2002a, 2002b) is representative of the widespread belief among practitioners that the "proper" KPI are related by cause-and-effect relations to measures of financial performance.

² Forrester (1994), summarizing the then mature field of systems dynamics, also has argued for the value of linked, systems models of performance.

briefly discuss these benefits, which include predictive ability, improved decision making, communication, learning, and goal congruence.

Predictive Ability. Cause-and-effect relations, by their nature, use leading indicators to predict key outcomes. If reliable, predictive relations exist in PMM, for example, leading measurements in non-financial areas can be used to predict future financial performance (Kaplan and Norton 1996: 8). Furthermore, analytical models demonstrate that the evaluation-weighting of measures can depend on their predictive ability (and decision sensitivity; e.g., Datar *et al.*, 2001). Goal-setting and expectancy theory research (Locke and Latham, 1990; Green, 1992) demonstrate that individuals are motivated to earn incentives when they believe that their efforts drive performance measures (and also when goals are achievable and rewards are based on measured performance). Multi-performance measure systems can be useful management controls, but they are not easily interpreted unless one can describe how a change in one criterion affects a change in another (Ridgway, 1956). Thus, if relations in PMM meet Hume's predictive-ability criterion for cause and- effect, they clearly can be useful to develop and control reliable planning scenarios.

Improved Decision-Making. Related benefits also can accrue inside the "black box" of predictive ability. Reliably predicting future effects of current actions and outcomes at key points in the value chain can aid decision-making (e.g., Eccles, 1991). Resource and capability-based strategy research predicts that superior decisions and performance will result from systemic management, rather than myopic focus on individual elements of the value chain (e.g., Huff and Jenkins, 2002; Sanchez *et al.*, 1996; Forrester, 1994). A PMM with valid, predictive relations is posited to reduce the cognitive complexity of both understanding and managing multiple measures of performance (Luft and Shields, 2002; Morecroft *et al.*, 2002). Furthermore, a predictive PMM can free managers to focus more on strategic and evaluation decisions than on information processing (e.g., Kaplan and Norton, 2001).

Communication. Cause-and-effect relations can enable effective communication of how best to achieve key operating and strategic performance. From a systems perspective, de Geus (1994) argues that even a simplified but credible PMM can be a powerful communication device. Magretta (2002) also argues that models to explain an organization's business activities are essential to tying strategic choices to financial results (see also Ittner and Larcker, 2001). Morecroft and Sterman (1994) further argue that PMM are effective when they become integral parts of management debate, dialogue, communication, and experimentation. Indeed, facilitating and communicating strategy via demonstrated cause and effect are some of the key "selling points" of Kaplan and Norton's (1996, 2001) balanced scorecard.

Learning. The cause-and-effect relations in a PMM demonstrate outcomes and tradeoffs among leading and lagging measures. Nonaka (1994) and Nonaka and Takeuchi (1995) argue that successful organizations institutionalize and perpetuate learning through creating, capturing, and communicating critical knowledge. PMM with cause-and-effect relations can educate managers and help them in controlling and committing to multiple measures (e.g., Feltham and Xie, 1994; Willard, 2005: 131).

Goal Congruence. Incentives based on single measures can induce incongruent behavior and management myopia (e.g., Ridgway, 1956; Dearden, 1969). Because a cause-and-effect PMM help individuals to see how their actions affect future performance, it fosters organizational focus and goal congruence (Kaplan & Norton, 2001, 2004). A strategy-driven PMM guides individuals to formulate local actions that contribute to achieving organizational-level strategic objectives. Hence, cause-and-effect relations direct managers' decisions to align the organization's limited resources with strategic outcomes.

Summary of Empirical Evidence for Expected Benefits

The beneficial effects of cause-and-effect relations allegedly support improved predictions, decision-making, communication, learning, and goal congruence. These outcomes should be observable in PMM users' strategic and operational choices and in operational and financial outcomes. Although influential literature clearly points to cause-and-effect relations as essential for the success of PMM, empirical support is minimal.

The few empirical studies of the existence or benefits of cause-and-effect relations in PMM are inconsistent. Contrary to Malina and Selto (2001), both Banker et al. (2000) and Ittner and Larcker (1998) find that relatively few managers and executives in their sampled firms had learned or understood any cause-and-effect relation between customer satisfaction and future profitability, although their incentive plans were linked to both. Lipe and Salterio (2002) find that experimental subjects made different but not necessarily better decisions related to alternative formats of performance measures (i.e., randomly arranged versus measures in displayed "balanced scorecard" categories). Ittner and Larcker (2003) observe that cause-and-effect relations among firms' multiple performance measures often are neither specified nor measured well. They find that companies rarely associate the actual impacts of changes in nonfinancial measures with future financial results. Bryant et al. (2004) associate cross-sectional data that proxy for outcome measures across four typical balanced scorecard perspectives to explain financial performance. In a more powerful test, Banker et al. (2000) use context-specific, time-series data to provide evidence on the impact of non-financial measures on firm performance. Neither of

the latter studies tests for cause and effect, but both document suggestive associations between customer satisfaction and future financial performance. Empirical evidence that supports the predictive ability of PMM has been in the form of uncritical self-reports (e.g., Rucci *et al.*, 1998). Indeed, most systems experts downplay the long-term predictive ability of complex systems models (e.g., de Geus, 1994). Hence, exploring evidence for the existence and benefits of cause-and-effect relations is the original motivation for this study.

RESEARCH SITE AND CAUSE-AND-EFFECT MODEL DEVELOPMENT

The host company for this study, a Fortune 500 firm, has sponsored two previous studies (Malina and Selto, 2001 and 2004).³ These earlier studies relied almost exclusively on qualitative analyses of extensive interviews with company and distribution managers. The findings of the two previous studies motivated the present study. In brief, the previous studies document that managers and distributors perceived that the DBSC:

1. Contains credible cause-and-effect relations among DBSC measures, although the company has neither expressed the relations as a “strategy map” nor conducted statistical testing of the relations
2. Communicates strategic intent effectively
3. Promotes goal congruence by effective communication and incentives to achieve strategic objectives
4. Directs distribution managers to change their processes and decisions to achieve DBSC targets
5. Failed to achieve the above when communication was ineffective, and
6. Has been revised repeatedly as the company seeks to include only accurate, reliable, and auditable DBSC measures

More recent interviews have disclosed that the company plans to deploy the DBSC to its global distribution network. The accumulated evidence leads the authors to believe that the DBSC is an example of an effective PMM that possesses qualities described in the normative BSC literature. The qualitative support for cause-and-effect relations in the DBSC (finding 1 above) is particularly motivating for this study. Thus, this study was initially motivated to answer what we believed was the only unanswered research question: Whether statistically reliable cause-and-effect relations actually exist in the DBSC. But we uncovered other interesting questions in the process.

³ See Malina and Selto (2001, 2004) for extensive descriptions of the research site, original interviews, and qualitative method used.

The DBSC Dataset

When the DBSC was introduced in the fourth quarter of 1997, it contained more than 20 key performance measures. After two years of evolution, the DBSC dropped to 11 somewhat different performance measures. A timeline of DBSC events is shown in Figure 1. Across the 31 quarters of data comprising this study (1997 – 2005), seven measures have been used continuously and have sufficient data for the statistical analyses that appear later in this study. Since Malina and Selto (2001), the company has reduced the number of distributors to 19 by merging lower performing units with higher performers. The 19 surviving distributors have up to 31 consecutive quarters of data. All available performance data are used in the analyses that follow. Table 1 contains the continuously used DBSC measures and brief definitions and explanations of the sources of the measures.

Insert Figure 1 and Table 1 here

DBSC Model Development

The company had not expressed its DBSC as a strategy map, which is a prominent feature of the balanced scorecard literature. We derived the DBSC map for this study from interview data using a method identical to the first method reported by Abernethy, et al (2005). The method analyzes the elicited knowledge of individuals within an organization first by coding interview transcripts for revealed performance constructs. We initially used Malina and Selto's 2001 coding of semi-structured interviews with five DBSC managers and nine distributors to determine the relations between pairs of measures in the DBSC that users and managers perceive. In total, 179 coded comments referred to variable relations.⁴ Following the PMM literature and our earlier work, we inferred cause-and-effect from interviewees' comments, and we initially coded 84 of these comments as cause-and-effect relations between specific pairs of variables.⁵ The summary of computer coding in row one of Table 2 generates the constructs or building blocks of the hypothesized cause-and-effect model.

Insert Table 2 here

The second step of the qualitative method to build a cause-and-effect map is to observe consistent patterns or relations among the coded constructs using relational queries in qualitative database software.⁶ Related constructs are connected with directional arrows, which we inferred from the nature of the

⁴ Interviewees discussed several other performance measures that at the time of this research did not have sufficient data to support statistical tests that are discussed later (e.g., service cycle time, which was believed to be a driver of customer satisfaction). For consistency with later statistical analyses, this study addresses the measures which were used for the entire time series.

⁵ Ninety-five additional comments referred to vague relations between one DBSC measure and other, unspecified drivers; e.g., "there are other measures that drive (financial measures)."

⁶ Ambrosini and Bowman (2002), Malina and Selto (2001), and Friese (1999) are among the studies that use the relational database feature of qualitative data software to build relational maps.

relation comments. In addition, we subjectively evaluated each relation for consistent expressions of relations rather than merely unrelated proximity. We validated this model by presenting it to two company managers, who were responsible for the administration of the DBSC and approved the model. Hence, we believe that we have properly specified the company's beliefs for cause-and-effect relations in the DBSC. Figure 2 is the visual representation of the DBSC.

Insert Figure 2 here

Figure 2 describes the cause-and-effect performance model that company personnel perceive as a map of organizational success. The time periods and extended boxes of Figure 2 reflect approximate temporal, lagged effects, which are posited to be integral to a PMM's cause-and-effect validity (H. Nørreklit, 2000). Managers and distributors expected time lags in the identified relations but could not be precise about the length of the lags. Distributors expect lags of one to two quarters, perhaps up to one year, for the effects of early value-chain performance measures (e.g., fill rate to customer satisfaction, customer satisfaction to sales growth).

TESTS OF CAUSE AND EFFECT

Despite widespread beliefs in cause-and-effect relations in PMM, statistical validation of causality is not trivial. Empirically verifying cause and effect requires effective experimental controls that rule out alternative explanations and permit cause-and-effect inferences. Clearly one cannot infer causality on the basis of covariation between variables. Although simultaneous cause and effect might exist, without careful controls one could not rule out that an unobserved variable was the cause of simultaneously observed effects. Time-series models of effects alone cannot provide evidence of causality; they test only for temporal precedence. Finally, predictive ability demonstrations are insufficient to support causality; they document out-of-sample regularities. A systematic, holistic approach is indicated. Hence, we employ a well-validated, rigorous econometric approach, Granger causality, to detect cause-and-effect relations in the DBSC.

Granger Causality

The fully developed concept of "Granger causality" (Granger, 1969, 1980; Ashley et al., 1980) is consistent with Hume's criteria and dominates testing for cause-and-effect evidence in economic models. The method proceeds in two steps. First, Granger causality is inferred from X to Y when significant correlation is observed between X and Y while considering *all available sources of information*. This condition supports or refutes the uniqueness of the relation, or alternative explanations. Operationalizing such tests literally is impossible in archival, quasi-experiments because "all available sources of

information” cannot be controlled or measured. However, tests of Granger causality customarily regress a dependent variable on lagged values of the dependent variable, Y , assuming that lagged values of Y and the hypothesized lagged independent variables capture “all available information.” Granger estimation tests support causality if coefficients of lagged independent variables, which capture time precedence, are significant as predicted in the presence of the lagged dependent variables (Darnell, 1994). Second, Ashley *et al.* (1980) propose that more rigorous Granger *mean* causality is inferred if the mean squared error of a forecast of Y is significantly less using a model of lagged X and Y (the full model) than using only lagged values of Y (the constrained model). If the full models have superior predictive ability, their root-mean-squared prediction errors (RMSE) and residual sums of squares (RSS) should be significantly smaller than those of the constrained models. Granger causality can measure theory, temporal ordering, high correlation, and predictive ability, which are the necessary elements of causality. The Granger tests we implement here (as in most archival studies) might support reliability, but only can refute cause-and-effect validity. This is consistent with most conventional notions of scientific inquiry that seek rejection of null hypotheses.

Hypothesized Cause-and-Effect DBSC Relations

The optimal lag structure of the DBSC is not apparent theoretically or from the interview data, but time-series models (not tabulated) indicates a consistently significant ($\alpha = 0.05$) one- and two-quarter lag structure in the DBSC’s dependent variables. We conservatively include dependent and independent variables lagged up to four quarters in the following tests to capture time precedence and “all available information.” The relations of the DBSC can be expressed as a system of linear path equations, which are derived from Figure 2:

$$(1) \text{PTO}_t = a_0 + \sum b_i \text{PTO}_{t-i} + \sum c_j \text{FR}_j + \varepsilon_t$$

$$(2) \text{CSAT}_t = d_0 + \sum e_i \text{CSAT}_{t-i} + \sum f_j \text{FR}_j + \gamma_t$$

$$(3) \text{WASG}_t = g_0 + \sum h_i \text{WASG}_{t-i} + \sum k_j \text{CSAT}_j + \delta_t$$

$$(4) \text{PBIT/S}_t = l_0 + \sum m_i \text{PBIT/S}_{t-i} + \sum n_j \text{WASG}_j + \sum o_j \text{PTO}_j + \sum p_j \text{WTO}_j + \sum q_j \text{SAFE}_j + \xi_t$$

where PTO is parts inventory turnover, FR is customer parts fill rate, CSAT is customer satisfaction, WASG is weighted average sales growth, PBIT/S is distributor profit before interest and taxes divided by sales, WTO is whole goods inventory turnover, SAFE is safety, and ε_t , γ_t , δ_t , and ξ_t are independent, normally distributed error terms.⁷ Right-hand-side summations ($\sum$) of the lagged dependent variables are from $i = t-1$ to $t-4$, and summations of the lagged independent variables are from $j = t$ to $t-4$. Granger

⁷ The regression residuals are not importantly ($r_{\max} = 0.055$) or significantly correlated ($\alpha = 0.05$) across equations, which permits the use of OLS (Bollen, 1989: 64, 404). Kolmogorov-Smirnov tests do not reject hypotheses that the prediction errors are normally distributed ($\alpha = 0.01$). These and other untabulated results are available from the authors.

causality tests require that the lagged independent variables are significant and, in this case, that all but one of the variable coefficients have positive signs, because of the nature of the posited relationships. The exceptions are coefficients “ q_j ” on $SAFE_j$, which are expected to be negative.

Quantitative Data

We originally had 14 quarters of data available to estimate the DBSC’s relations and quarters 15 – 17 to use as a hold-out sample to test the DBSC’s predictive ability. The initial tests of multiple, alternative specifications were unsupportive of causality in the DBSC (see Table 3), with only one statistically significant, hypothesized cause-and-effect relation, which indicates that sales growth, lagged four quarters, might cause distributor profitability in a linear Granger model. However, all of the tested relations have uniformly inferior predictive ability (not tabulated), refuting Granger causality. This evidence points to a noisy model that a successful firm clings to for no apparent good reason.

Insert Table 3 here

Since the time of the initial analysis, the company has refined both measures and measurement methods to improve the accuracy and verifiability of DBSC performance (Malina and Selto, 2004). Therefore, we have reason to believe that analysis of an expanded and improved dataset that is now available might reveal the expected cause-and-effect relations among DBSC measures. The expanded dataset includes 31 quarters of DBSC data (1997 – 2005), which include the 17 quarters used initially. Analogous to the initial study, we use twenty-eight quarters of data (Q1 – Q28) to estimate the DBSC relations and the remaining three (Q29 – Q31) as a prediction sample. Since the initial analysis, nine of the 31 distributors were merged with larger and better performing distributors by the third quarter of 2004 (quarter 28); two others were merged one quarter later; one was merged two quarters after that, leaving 19 distributorships. All available data are used to estimate the each of the DBSC relations because the expected relations should apply to all distributors, regardless of performance or merger status.⁸

Descriptive statistics and pair-wise correlations for the estimation set of the expanded (un-lagged) data are presented in Tables 4 and 5, respectively.⁹ Exploratory factor analysis of the seven (un-lagged) DBSC variables simultaneously indicates that further data reduction is not necessary (results not tabulated). Correlations in Table 5 are generally small and indicate lack of multicollinearity.

⁸ We are unable to cleanly analyze only the 19 continuing distributorships for the entire time-series because post-merger data is consolidated. Analyzing data for only the 19 survivors during the pre-merger time period, 1997 Q1 – 2004 Q3, generates results that are less favorable to Granger causality than reported here.

⁹ Initial descriptive analysis shows that WASG has a large range for a proportional measure. Further investigation reveals that two distributors entered new markets early in the time series and had exceptionally large percentage sales growth in those markets in the first year, growing from a near-zero base. All reported results retain the five outlying observations of WASG from these two distributors; omitting these observations slightly improves the significance of several tests involving WASG, but does not affect results for other tests.

Insert Tables 4 and 5 here

Granger Estimation Tests Using the Full Dataset

Column two of panels A, B, C and D of Table 6 presents the linear Granger test results of DBSC relations for the full dataset. These results show improvement over the initial dataset, with some statistically significant relations among DBSC measures in the predicted directions. Parts fill rate (FR) in panel A, column 2 does not cause parts turnover (PTO), however. In panel B, fill rate (FR) has a statistically significant, contemporaneous *association* with customer satisfaction (CSAT), as believed by company personnel ($p < .01$), but no lagged effects that support cause and effect. In panel C, customer satisfaction (CSAT) does not cause sales growth (WASG). In panel D, contemporaneous sales growth (WASG) and parts (PTO) are *associated* with distributor profitability (PBIT/S), as believed ($p < .01$), but these associate current variable values and do not support causality. However, the four-quarter lag of sales growth (WASG4) appears to cause distributor profitability ($p < .001$).¹⁰ Therefore, the Granger estimation tests indicate a possible cause-and-effect link in the DBSC: distributor profitability, PBIT/S, might be caused by one-year lagged sales growth, WASG4.¹¹

Insert Table 6 here

Granger Predictive Ability Tests Using the Full Dataset

If the full models have superior predictive ability, their RMSEs and RSSs should be significantly smaller than those of the constrained models; that is, all percentage differences in Table 7 should be significantly negative and all F-statistics should exceed critical values. The predictive ability results were prepared as follows:

1. Estimate each of the four out-of-sample outcomes (PTO_t , $CSAT_t$, $WASG_t$, and $PBIT/S_t$) using the full, estimated equations in table 6 (including all hypothesized lagged variables).
2. Estimate each of the four out-of-sample outcomes using constrained equations that contain only the lagged *dependent* variables (estimated equations not shown), which provide predictive ability benchmarks.
3. Compute and compare RMSEs and RSSs across the pairs of equations for each dependent variable observation.

¹⁰ Chow F-tests using identical datasets show that three of the four full models (explaining CSAT, WASG, and PBIT/S) have statistically superior explanation ($p < 0.05$) compared to constrained models. However, only one of the improvements in full-model explanations is driven by a lagged driver ($WASG4 \rightarrow PBIT/S$).

¹¹ Estimations of relations omitting the first six quarters, which encompass almost all missing data, are not significantly or materially different from the results reported here.

We use the last three quarters of available performance data (quarters 29 – 31) to test the predictive ability of the DBSC equations estimated with the earlier 28 quarters' data. Table 7 shows that two relations show worse predictive ability with higher RMSEs and RSSs (dependent variables = PTO and WASG). In contrast, the full equations to explain customer satisfaction, CSAT, and distributor profitability, PBIT/S, do have 2 and 3.5 percent better predictive ability than the respective, constrained counterparts.¹² The better predictive ability of the full PBIT/S model, which is an out-of-sample test, indicates that the estimation results for that model are not sample specific, at least with regard to the impact of sales growth. However, neither improvement in predictive ability is even marginally statistically significant by F-tests (Johnston, 1994: 505) of differences in RSSs ($\alpha = .1$). Predictive ability and estimation results offer weak support of causality in the PBIT/S equation (4). No evidence supports causality in the other three DBSC equations or for other hypothesized causes of distributor profitability (PTO, WTO or SAFE).

Insert Table 7 here

Alternative Models

We also investigate alternative specifications of performance relations. Columns three through six of Table 6 display the estimation results for four alternatives for each DBSC relation.¹³ In column three are the results of Granger causality tests including fixed distributor effects, which are binary (0, 1) variables. We include distributor effects in an attempt to capture more of the set of “all available information” and because each distributor might face different market conditions or exert different efforts. Some of these binary variables are highly significant, but most inferences about variables of interest are no more favorable to Granger causality than in equations without these effects. The only exception in column three is a significant relation between SAFE4 and PBIT/S ($p < .05$). The negative sign suggests that lost-time accidents, lagged by four quarters, negatively affect distributor profitability, perhaps through increased insurance costs. We caution that this is an isolated result.

Column four reports nonlinear (natural log) transformations of equations (1) to (4), without fixed effects. The nonlinear specification of the WASG model (omitting negative sales growth observations) indicates a highly significant four-quarter lagged effect of customer satisfaction (CSAT4) on sales growth ($p < .01$). Although it is later than company personnel expected, this nonlinear result is consistent with prior research by Ittner and Larcker (1998) who suggest that distributors might need to exceed customer service thresholds to impact sales. Because we cannot test for explicit threshold effects, we regard this nonlinear result with caution. A result in panel D shows another significant, nonlinear relation between a

¹² The estimation and predictive ability tests were repeated alternatively holding out 1, 2 or 4 quarters with nearly identical results.

¹³ No predictive ability inferences were materially different from the results reported earlier.

one-period lagged effect of sales growth (WASG1) and distributor profitability (PBIT/S), but this is a weaker effect ($p < .05$) than found from a four-period lag in the linear Granger models. Note that FR4's significant, nonlinear result in panel A is incorrectly signed.

Finally, columns five and six report results of one-quarter and four-quarter differences or changes models. Changes models can control for distributor-level market and effort effects that might be masked in the original Granger specifications. The 4-quarter PTO model in panel A shows a significant effect of a corresponding change in FR ($p < .05$). Similarly, the 1- and 4-quarter changes models of CSAT in panel B show significant impacts of corresponding FR changes ($p < .05$, $p < .001$, respectively), but contemporaneous FR also is significant in other model specifications. The 4-quarter change in PBIT/S model in panel D shows highly significant effects of corresponding changes in WASG ($p < .01$) and SAFE ($p < .01$). The 1-quarter PBIT/S change model finds a significant relation with WTO ($p < .05$). These observed effects are inconsistent, but more importantly, they are ambiguous about causality because they associate contemporaneous changes and do not cleanly establish time precedence.

Summary of Granger Tests and Additional Considerations

In summary, the time-series data provide some support for cause-and-effect relations in the DBSC, but the case is inconsistent and not compelling. Several lagged, independent variables are significant “in the presence of all other information,” but most are not. Predictive ability is not established consistently and never significantly. We find one significant customer satisfaction relation in a nonlinear model between a four-quarter lagged effect of customer satisfaction (CSAT4) and sales growth (WASG). The only support for cause and effect across multiple model specifications appears in a four-quarter lagged effect of sales growth (WASG4) on distributor profitability (PBIT/S) in several linear Granger models. However, this statistical significance in the model is accompanied by insignificantly improved predictive ability. The case for cause and effect in the DBSC overall is quite limited.

Distributors' performance on DBSC measures exhibit signs of conformity to company targets. Figure 3A presents the time-series of proportions of distributors' target performances for CSAT. The proportions of Red, Yellow, and Green distributors obviously shift to Green performance over time. We also visually examined regression model residuals for evidence of the development of expected relations over time. Figure 3B presents an error-bar chart of the performance time series for customer satisfaction (CSAT) residuals from a linear regression of equation (2) using the full 31 quarters of data. The early time-series of CSAT regression residuals is noisy but reflects obvious tightening and overall improvement in the last

five quarters. These results show that the DBSC is related to impacts on performance despite lack of demonstrable cause-and-effect.¹⁴ Figures for other performance measures are similar.

Insert Figure 3 here

Our multiple tests find that the DBSC has limited significance and predictive ability, which refute cause-and-effect relations in the DBSC as an explanation for its continued use. This apparently flawed model could preclude reliable prediction, decision making, learning and communication. Yet distributors' DBSC performance has improved, and the company has continued to use the DBSC in subsequent periods. In fact, the company has placed more weight on the DBSC for organizational change and for variable compensation of distributors and is deploying the DBSC to its worldwide distribution channel.

Finding evidence to support cause and effect within PMM might be possible, but not easily; furthermore, such evidence conservatively only can refute cause and effect (Popper, 1959, 1963). In the context of PMM at least three reasons work against establishing Granger causality: (1) managers adapt the firm's actions and the underlying production function to PMM and other feedback (hence, statistics are unstable); (2) a PMM that is not a fully specified input-output model may not reflect underlying cause and effect sufficiently; and (3) cause and effect might not exist in non-physical (portions of) PMM (e.g., relations of service performance). This study's empirical findings, which are either contrary to normative theory or reflect a PMM that cannot exhibit cause and effect, motivate our continuing the study. In the case of the enduring DBSC, explanations other than cause and effect are required. Hence, we believe at least two theoretical explanations exist for why a PMM can endure without evidence of cause and effect relations: (1) misspecification of DBSC relation types and (2) an incomplete theoretical framework.

RECONCEPTUALIZED THEORETICAL FRAMEWORK

As amply discussed previously, the DBSC does not pass rigorous Granger causality estimation and predictive ability tests, but the DBSC is an enduring PMM at a successful company. We do not accept that managers' beliefs about the DBSC are either irrational or deceptive, and our failure to find evidence of cause and effect challenges our original beliefs about whether cause-and-effect exist or are necessary in the DBSC. These results lead us to reconsider the nature of the relations in the DBSC that were previously published (Malina and Selto 2001, 2004). Other types of relations can and likely do exist in the DBSC and probably in other PMM. An expanded analysis of relations made us realize that logical and

¹⁴ At the suggestion of a reviewer, we investigated "distortion" in the company's performance targets (Baker 2002); that is, we test whether distributors' performance ratings (red, yellow, green) are consistent with profitability. We regressed quarterly distributor profitability (PBIT/S) on the contemporaneous number of red, yellow, and green ratings received on DBSC measures. The results show negative associations with red ($p < .001$) and yellow ($p = .192$) ratings and positive associations with green ratings ($p = .004$). These results indicate no performance target distortions that might explain lack of observed cause and effect between DBSC measures.

finality relations also can exist among measures in PMM. Importantly, these other relations are consistent with managers' continued use of the DBSC for management control. The classifications of relations reflect more than semantic differences; the differences have important implications for PMM development, validation, use, and feedback. We discuss logical and finality relations, then we re-analyze our DBSC results.

Logical Relations

Logical relations exist by human construction or definition, and may be common elements of PMM. They are the results of related human constructs, such as mathematics, language, and accounting (L. Nørreklit, 1987: 164; Ijiri, 1978, ch. 4 & 5). Logic, for example, defines that debits equal credits, and in general logic is a consistent tool for creating and managing human reality. Financial and management accounting systems, DuPont models (ROI), and net-present-value calculations are common examples of logical models that measure economic profitability. Although specific applications often vary, the logical relations of these models are independent of firm-level contingencies. In accounting the effect of an action on profit, (e.g., sale of a product with a positive contribution margin), necessarily occurs by the double-entry logic of the accounting system, not cause and effect. Note that the relation between two phenomena cannot be both logical and causal. In a cause-and-effect relationship, the cause happens before and independently of the effect, and the cause must be logically independent of the effect.

It is a social fact [Searle, 1995] that financial accounting models are used in our society to measure and evaluate the financial performance of a firm. This implies that financial analysis is needed in a firm to structure and evaluate the economic aspects of decisions and actions. For example, the creation of profitability through making customers loyal depends on the revenues and costs of making them loyal, which dictates that we have to use financial analysis to evaluate whether a loyal customer is profitable (H. Nørreklit, 2000). Therefore, any financial logic embedded in the PMM has to be linked to the rules of financial accounting performance (H. Nørreklit et al. forthcoming), not cause and effect.

Logical models, such as accounting, are not refutable by empirical evidence, only by deductive reasoning. For example, decomposed DuPont relations, such as ROI equals return on sales multiplied by asset turnover, are logical, not cause-and-effect relations. A regression model of these logically related variables does not generate empirical evidence on the validity of the logic or formula. Statistical significance, or the lack thereof, speaks instead to the reliability of *ceteris paribus* conditions that support the logical relation, such as control of pricing and costs and other related but omitted logical links. In practice, many activities logically influence profitability, but PMM appear to be simplified combinations of KPI, not fully specified accounting models. Inevitably, logical *ceteris paribus* conditions will be

difficult to control or observe in actual PMM, and statistical explanation of relations among logical KPI will be less than perfect, perhaps insignificant.

Finality Relations

A finality relation exists when (a) one believes that a given action is the best or most desired means to an end, and (b) the belief, desire, action and end are related by custom, policy, or values (Arbnor and Bjerke, 1997). Actions driven by finality are performed because the actions conform to the beliefs and wishes of a person (or group). Acceptable outcomes (e.g., profitability) can reinforce these finality relations, but can not transform finality into cause and effect. Finality is fundamentally different from cause and effect because finality-driven actions and outcomes are not independent or uniquely observable (Mattessich, 1995). They are confounded and violate Hume's first criterion of independence of phenomena. Furthermore, observation of subsequent favorable outcomes reflects the results of an engineered process, but does not signal a generic process to that end.

Finality relations have other characteristics that set them apart from cause and effect. Unlike cause and effect, any chosen means is but one of several or many, which may be used to reach the end. Furthermore, a finality relation can be idiosyncratic to a particular setting or context (Arbnor and Bjerke, 1997: 176). For example, corporate vision and mission statements commonly contain finality relations. Contingency theory is a common academic expression of finality in organizations, and empirical tests of contingency concepts often do not generalize beyond a specific firm or sample (Van de Ven and Drazin, 1985; Chenhall, 2003). The oft-cited role of a BSC to tell the "story of the company's success" is another expression of finality that directs employees to preferred actions that might not be generalizable beyond the specific company and time. Finality relations often rely on incomplete arguments where premises are lacking, such as unspoken, *ceteris paribus* conditions that are nearly impossible to control in natural settings. This complexity of relations, in conjunction with lack of independence of phenomena, is an indication of finality rather than causality. However, to use finality relations to achieve sustained control of actions, a finality belief that a given action leads to an end must be reliable or perceived as such, at least in a specific context. In many practical situations of management control, finality and logical relations work tightly together when one uses financial analysis to decide on strategies and policies, similar to Simons' (2000: 276) belief-system controls. For example, *ceteris paribus* conditions that exclude unprofitable products and customers might engineer a reliable relation between customer satisfaction and profitability.

Statistical analysis might be helpful to establish context-specific reliability of a finality relation, but it cannot be definitive. Validating a finality relation as the best or unique means to an end is complicated by equifinality and finite data. Although statistical validation may not be possible, financial analysis of costs

and benefits of finality-driven strategies might explain their use and longevity despite statistical insignificance of finality relations among measures.

In summary, statistical tests, such as Granger causality, are appropriate for validating cause-and-effect relations, which may be uncommon except in PMM that reflect physical productive processes. In contrast, statistical tests are irrelevant for establishing the validity of logical relations and may be insufficient for finality relations, both of which may dominate most PMM. Furthermore, feedback from financial analysis of logical and finality relations may explain the duration and evolution of PMM to a greater degree than statistical analysis. Thus, our general lack of statistical support for the enduring DBSC reflects the presence of logical and finality relations among financial and non-financial performance measures. Company support for the DBSC may reflect its favorable impact on company profit, which results from a tangled chain of financial logic and finality that is not observable at the distributorship level and might be exceedingly difficult to discern at the company level.

Re-analysis of the Data from Model to Results

Our previous beliefs about cause and effect in the DBSC were based on normative assumptions and qualitative analysis, which, like other empirical methods, is subject to researcher bias. Hence, we expected cause-and-effect relations, and we found them. However, the statistical results and our more refined understanding of relation types in PMM challenge those prior beliefs, which were reinforced by the original qualitative analysis. If we had approached the qualitative data with broader, less dogmatic beliefs about the nature of PMM relations, perhaps we would have reached different conclusions about the prevalence cause-and-effect in the DBSC and the applicability of Granger causality tests to this case.¹⁵

With a wider theoretical lens, we recoded the original and 2005 qualitative interview data by asking:

1. Does this relation reflect independence of phenomena, time precedence, and predictive ability?
2. If so, code the relation as cause and effect.
3. If not, code the relation as logical or finality, as appropriate.

Re-analysis of the qualitative data reveals no unambiguous, cause-and-effect relations, often because of violations to the independence criterion. The DBSC's logical relations of financial cost-benefit are now obvious, but we had coded them previously as cause-and-effect. The DBSC contains an inventory replenishment relation (equation 1) plus familiar relations between inventory turnover and distributor profitability (equation 4), and relations between revenue and cost drivers and profitability (equation 4). All are logical relations, not cause and effect, because of their derivation from the accounting system. Similarly, we now identify the customer satisfaction (CSAT) relation with fill rate (FR) as a finality

¹⁵ This is a major point of "grounded theory" approaches to qualitative research (e.g., O'Connor et al., 2003; Dougherty, 2002; Corbin and Strauss, 1990).

relation because having parts available on time is the company's preferred action to increase customer satisfaction, but that is hardly the only approach. Likewise, the relation of customer satisfaction (CSAT) driving sales growth (WASG) most likely is finality, not cause and effect because of the numerous ceteris paribus conditions required. We now classify all DBSC relations as logical or finality, and none as cause and effect, as shown in Table 8.

Insert Table 8 here

Revisit the four DBSC equations, which are abbreviated below. Other variables (generically symbolized by Z_t) are proxies for "all available information" and, as before, are not of direct interest to this study.

$$(1) PTO_t = f(FR_t, Z_t)$$

$$(2) CSAT_t = g(\mathbf{FR}_t, Z_t)$$

$$(3) WASG_t = h(\mathbf{CSAT}_t, Z_t)$$

$$(4) PBIT/S_t = k(WASG_t, PTO_t, WTO_t, \mathbf{SAFE}_t, Z_t)$$

Italicized variables represent five logical relations; bold variables represent two finality relations. We explain and illustrate our revisions more fully as follows.

Logical Relations. The relation of parts turnover (PTO) as a function of fill rate to customers (FR) (equation 1) is a relation that derives from the logic of inventory replenishment, but other ceteris paribus conditions surround this logical relation. For example, the expressed uncertainty about fill rates from the company can induce distributors to build inventory levels to insure favorable fill rates to customers. This problem was identified by most distributors. Consider several distributors' explanations:

As we are customers of the factory, (fill rate) is very important to us. If we aren't receiving a high fill rate from the factory, we can't achieve a high fill rate to our customers. It's a domino effect. The factory is having availability problems now... If one piece of the (distribution) channel breaks down, all the pieces are greatly affected. (Distributor D)

What about our fill rate from (the company)? Big interaction there... Our fill rates are always higher than theirs. Theirs is 61% to us and ours is 90% to customers. We have to stock more inventory than them. (Distributor H)

Distributor E explains the logical impact of inventory turnover on distributor profitability (equation 4).

Obviously if you have less inventory and you still have good availability (of parts to customers) then you'll have more cash available, and less expense which will make you more profitable.

(Distributor E)

The logical impact of safety on distributor profitability (equation 4) also is evident in comments, such as:

It's more costly after the fact than it would be to build it (safety) into the process and show where it fits into the cost of doing business. If you look at workers compensation cost, the cost of medical care today, and injury intervention, all of those things come off the profit side of the business.

(Manager N)

The relation between weighted average sales growth (WASG) and increase in distributor profit (PBIT/S, equation 4) likewise is a logical relation of financial cost-benefit. The results in panel D of Table 6 show a consistently significant logical relation with current weighted average sales growth ($p < .001$) and with a four-quarter lag ($p < .001$).¹⁶ The consistent significance indicates that this relation must be tightly controlled; that is, increased sales of *profitable* products to *profitable* customers drive profits when key ceteris paribus conditions are maintained. A recent interview with a senior executive of the company confirms that the company controls conditions in the relations between sales and profit. Top management has decided which products are most profitable (to the company) for a distributor to sell, and it limits distributors' profitability by setting minimum product price markups, which appear to hold. The lagged effect means that it can take approximately one year for an average new customer to become profitable to the distributor. Although customers might be profitable immediately to the company, because of the company's control of products and prices, the costs of extra services and customer development borne by the distributor appear to not pay off quickly. Distributors, of course, recognize that they must absorb these costs.

They have not given us any tools to sell the product over the competition. This is a price sensitive market, and we're holding the line on our prices, and we're not giving away incentives like our competitors. They need to adjust the (sales) target if they aren't going to help us. (Distributor A)
(The company) is not worrying about what it is costing distributors to improve. They are looking at their cost. (Distributor G)

Finality Relations. We recoded some relations as finality rather than cause and effect. For example, the commonly voiced argument which follows describes a complex finality relation that involves achieving high first-time parts fill rate (FR) to customers as one way to improve customer satisfaction (CSAT, equation 2) and which must require many controls to be valid.

The measure (FR) is important and quite valid It is a direct measure of how well we serve our customers. If we are doing 99 percent, we are only disappointing 1 percent of the customers. It is a valid measure because it tells us how we are doing in giving the customer what they ask for the first time. People are very sensitive. They let us know if we're not living up to expectations. Some

¹⁶ The logic of the significant, nonlinear, one-period lag effect instead in column 4 is not intuitively obvious.

of our dealers are looking elsewhere to get parts because of the stocking (fill rate) problem.

(Distributor A)

Clearly, the fill rate is an important measure, but the relation to customer satisfaction cannot be cause and effect. A consistently significant contemporaneous result in panel B of Table 6 ($p < .01$) does support respondents' strong finality belief that a higher fill rate is associated with higher customer satisfaction. However, it is impossible to determine if other factors not included in the PMM, including other dimensions of service quality, are driving this result. The relation also appears to be idiosyncratic to this company's preferred approach, because other means to improve customer satisfaction surely exist (e.g., lower prices, fewer processing mistakes).

A finality relation also exists between customer satisfaction (CSAT) and sales growth (WASG, equation 3). At the operational level, increased customer satisfaction is not free but may increase sales. Sales growth also can be affected by uncertain factors that might not be controllable by distributors. These include competitors' actions, industry changes, and changes in customer values and tastes. The results in panel C of Table 6 indicate that this is not a reliable finality relation, because only one significant result is found across the five model specifications (one of 14 coefficients).¹⁷ Either customer satisfaction as a driver of sales growth is an invalid belief or macroeconomic factors like sales prices or industry effects influence the relation but are not controlled.

In light of the re-classified DBSC relations, the weak Granger causality results reported earlier are no longer surprising. Logical relations cannot be validated or invalidated by the statistical tests. The DBSC's far from perfect R^2 can be attributed to lack of control of important *ceteris paribus* conditions, such as distributors' response to the company's fill rate. The finality relations have incomplete reliability, which may reflect a dynamic environment and adaptive behaviors.

PMM AND THE CLIMATE OF CONTROL

Cause-and-effect relations among measures appear important for prediction, decision making, communication, learning, and goal congruence. Certainly cause and effect might exist in some PMM. In this case, however, the company's reliance on the statistically weak DBSC lead us to consider whether cause and effect are necessary to the success of a PMM. The reinterpretation of DBSC relations as logical and finality relations does give us better understanding of the weak statistical results obtained, but it does not by itself provide a convincing argument for why the company continues to support the DBSC and

¹⁷ Intrigued by this result, we also estimate the lagged, nonlinear (log) CSAT $\rightarrow$ WASG model with up to 28 observations and repeated the estimation and predictive ability tests. These tests are hampered by the need to omit negative sales growth values, but CSAT4 is significant ($p < .05$). Predictive ability, while better than a constrained model by 1.26 percent, was not significantly improved. Thus, even censored data that are most favorable to causality refute cause and effect.

plans to expand its use, and why distributors increasingly conform to performance targets. It is possible, for example, that the threat of the loss of the distributorship contract is sufficient to coerce conforming behavior. However, voluntary turnover other than retirement among the distributors is almost unheard of at this company, and most distributors agree with the intent of the DBSC (Malina and Selto, 2001). Both indicate a mutually beneficial relationship. We posit that an organization, like the one studied here, may use a PMM to reflect and reinforce a “climate of control” that reflects the company’s environment, style of management and institutional and social cultures. Furthermore, the climate of control achieved and reflected in financial success explains the longevity of PMM,

Contingency research in management control (e.g., Abernethy and Lillis, 2001) recognizes the importance of “fit” between strategy, culture, management style, uncertainty, and performance measurement as important to the design and effectiveness of control systems. Thus, we posit that *intended* fit influences the design of PMM such as the DBSC. We further posit that beliefs about relations among strategically important variables influence managers to create PMM with logical and finality relations that are supported by financial feedback and cause-and-effect relations, which may be validated by statistical tests. Uncertainties about these relations may contain much of the uncertainty construct that contingency research often measures poorly. A firm may install a PMM that reflects its climate of control to communicate its strategy to enhance learning and the legitimacy and fairness of goals and performance measurement. The aim of this PMM-based communication would be to increase motivation and to improve decisions and financial success (e.g., Anthony and Govindarajan, 1998: 7 & 95). We reason, therefore, that a firm might regard a PMM as effective if it contributes to goal congruence and desired conforming behavior, reinforced by improved financial performance. For example, top management at the company confirms that a control environment of *ceteris paribus* conditions for sales and profitability are maintained in the company. We posit that the DBSC enhances the company’s climate, which conspicuously features pay-for-performance and result control (Malina and Selto, 2001, 2004). Furthermore, the DBSC affects motivation and conformity favorably if the means of performance measurement are regarded as legitimate and fair. These elements of climate of control are illustrated in Figure 4. We next discuss the elements of our proposed theory of PMM effectiveness in the context of the DBSC.

Insert Figure 4 here

Pay for Performance Culture. As discussed earlier, distributors appear to have ample reasons to regard the DBSC measures seriously. Both variable compensation (now about 50% of total compensation) and contract renewal depend on DBSC performance. The DBSC is the foundation for the pay-for-performance climate created for the distribution channel. A few DBSC measures are controllable (e.g., safety) by

distributors, but many are influenced by less controllable factors. For example and as described previously, a distributor's parts fill rate to customers and its inventory policy are affected by the company's parts fill rate to the distributor. Therefore, the company uses the DBSC for relative performance evaluation (RPE) by ranking and comparing distributors by DBSC performance. The pay-for-performance effects on motivation are readily apparent in these representative statements from the interview data.

No one wants to be #31. They are very competitive people. (Manager L)

If a distributor is in the bottom quartile for 2-3 quarters in a row, then they are on probation.

(Manager J)

We are competitive. Anytime you publish a report and there 31 entities being measured using the same metric, it matters what rank you are. Even if no one looks at the rank, I want to be #1.

(Distributor E)

Results-Oriented Control System. Interview data show that the DBSC was designed as a result control. Merchant's (1998) four conditions for effective result control are: knowledge of the result desired, controllability of the desired result, measurability of the controllable result, and performance targets. Although controllability varies across measures, the DBSC addresses these conditions and focuses distributors on results that benefit the company. Consider the following selections from many similar quotations:

(The DBSC) is a way to measure them (distributors) in a balanced way, what they are really responsible and accountable for. It provides appropriate weights for what you want them to do. The company will benefit because you have attached certain weights that you want to drive them to perform well in. The weights make them swing that way. It's driving behavior toward the higher weights. (Manager K)

It's extremely and painfully obvious which are the most important (results). If you're the worst in (X) market share, you can't overcome it by greens in other areas. That's the lifeblood of the company. (Manager L)

When (the company) added new measures that they didn't tell us about and then they were red, it's not a subtle sign that we need to look at that area. (Distributor F)

Legitimacy. Although the company did not use outside consultants, it prominently named its model the "Distributor Balanced Scorecard" and used Harvard Business School-educated employees to design it. Although some distributors regarded the DBSC warily at first – especially noting its early lack of "balance" – none openly challenged more than small parts of it. Even if the PMM is always a "work in

process,” an organization can use PMM to build legitimacy by projecting rationality and efficiency to internal and external constituents (Carruthers, 1995; Meyer and Rowan, 1977). Most distributors accepted the model and its norms as legitimate. A common sentiment was:

I like all A's on my report card, so I want all of them green. I agree with almost all the measures.

They are indicative of where you are. (Distributor F)

Interestingly, neither the distributors nor the company had conducted statistical analyses to validate the DBSC. However, they observed relations between performance measures, such as that between fill rate and customer satisfaction. This reinforcement of beliefs also adds to the legitimacy of the DBSC. For example, consider this almost unanimously expressed belief:

(Parts fill rate) measures whether we have the right type of inventory parts on hand and the right quantities. It's one of the most important measurements we have here. The key thing is the right product mix and quantity and to satisfy the customer the first time around. (Distributor D)

Fairness. Prior to the DBSC, managers and distributors acknowledged that subjectivity and favoritism affected management of the distribution channel. The DBSC was intended to make evaluations and evaluation processes appear more fair and objective (e.g., Burney et al., 2006). Managers may accept a PMM if it persuasively builds on the ideas of the market economy and “fair contracts,” which govern social relationships in the US (Bourguignon et al., 2004). The idea of fairness expresses the opportunity open to everyone to work their way from the bottom to the top. Everyone is expected to act freely under contracts to which s/he chooses to be committed and under a general moral claim to fairness. Furthermore, fairness is associated with suitable remuneration for a person's work performance and with the equal treatment of everyone (d'Iribarne, 1994). Consider the following representative quotations from distributors:

(The DBSC) is intended to be a way that the factory can measure the performance of distributor network in such a manner that it puts everyone on a level playing field as far as measures.

(Distributor D)

As (the company) did the every-3-years contract review, I had heard that there was speculation that some guys got an easier or harder approach based on whether they were friends or enemies of (the company). (The DBSC) at least gave some quantitative basis to the evaluation process. It's more objective and black and white on key areas. (Distributor F)

I grew up working for a CPA and he ingrained in me that if you can't measure it, you can't improve it. I like this because its measures I already have and because it takes some of the guessing out of “how does (the company) view me?” I just like knowing my grades. I assume that if I have a

green, (the company) is grinning. (The DBSC) helps me think that greens will take the stress away for the next contract review. (Distributor F)

Motivation and Conformity. On the basis of getting what one measures and rewards (i.e., Merchant, 1998), one expects that distributors will manage their processes (and perhaps the measures) to achieve favorable performance ratings. Even in the absence of demonstrated reliable measure relations, acceptance of the system and conforming behavior also would be consistent with a DBSC that has a primary purpose to create a climate of control through pay for performance, fairness, and legitimacy.

Archival performance data shows conformity to norms over time for many DBSC measures. The company has merged 11 distributors over the past two years based on DBSC results and a desire for better overall distribution efficiency. Although the mergers have created periods of performance instability, general improvements in DBSC results for most measures are apparent. For example, the percentages of distributors attaining the green (highest) scores on heavily weighted customer satisfaction have increased over time, while those distributors with yellow and red scores have decreased over time (see Figure 3).

Climate of Control Summary. The model of PMM effectiveness that we propose is in Figure 4.¹⁸ Although cause-and-effect relations seem desirable, they may be unnecessary or infeasible in a highly uncertain, dynamic environment. Even in stable conditions when cause-and-effect relations are indicated, as a practical matter any observed lack of statistical reliability may be attributed to a PMM's continuing evolution. As long as the organization is committed to achieving a reliable PMM in the future, a PMM could be an otherwise effective control device despite its current lack of statistically reliable relations among measures. More often, perhaps, PMM will contain logical and finality relations that can support the desired climate of control. By designing a PMM to be a result control and pay-for-performance tool, and by establishing its fairness and legitimacy, management can motivate employees to conform to company expectations. Thus, the climate of control might be sustainable even when performance relations cannot be unambiguously or statistically demonstrated as cause and effect. This climatic role for PMM might outweigh a PMM's usefulness for prediction and decision support. In an uncertain environment, the rhetoric of a balanced scorecard model combined with face-valid measures, valid logical relations, credible finality relations, and positive financial feedback may be sufficient for a PMM to be considered successful.

¹⁸ We gratefully acknowledge a reviewer's constructive comments to improve this figure.

CONCLUSIONS, LIMITATIONS AND FUTURE RESEARCH

Conclusions

Cause-and-effect relations among performance measures have been argued to be essential features of performance measurement models (PMM) because they can aid financial prediction and decision making as well as create effective learning, communication, and goal congruence. We approached this study with the intention of testing the validity of cause-and-effect relations in an enduring PMM at a *Fortune* 500 company. The company established a distributor balanced scorecard (DBSC) for its distribution channel. Qualitative data from interviews with managers and distributors prior to statistical tests are reflected in perceived relations among the DBSC measures, a finding that establishes face validity for the model tested statistically (see Malina and Selto 2001, 2004). We evaluate the DBSC for evidence of Granger causality, but find at best limited support for any cause-and-effect relations both in initial and expanded time-series datasets. Our statistical results pointed to explanations that we could not accept – that the DBSC must be a fad or a deceptive exercise of management power – because the DBSC has endured and worldwide deployment is planned. Statistically unreliable relations thus far have not been a barrier to continued and more confident use of the DBSC in the North American distribution channel of this large, successful, international firm.

This dissonance motivates a review of the types of relations that can appear in PMM, and this broader review identifies two other types of relations in PMM, logical and finality relations, that can complement or might supplant cause-and-effect relations. Without a proper understanding of the different types of relationship, a deeper understanding of the design and use of PMM might not be possible. For example, any PMM relation involving financial measures of performance reflects accounting logic that cannot be refuted by empirical evidence. The different relations combined with a further analysis of both qualitative and quantitative data lead us to conclude that cause-and-effect validity might be less important to some contexts than a PMM that is perceived to be legitimate and fair and that supports an effective climate of control. Our careful use of theory both to motivate the cause-and-effect study and to interpret the results indicate that justifying PMM only on the basis of valid cause-and-effect appears to be myopic in this case. Hence, this study indicates that one should not reject the validity of a PMM simply because statistical evidence of cause and effect is lacking. Organizational validity may lie elsewhere, as summarized in Figure 4. Whether and when a PMM successfully supports an effective climate of control without intended or validated cause-and-effect relations deserves future research.

Previous studies (e.g., Malina and Selto, 2001) have concluded that PMM can be effective strategy communication and motivation tools. The present study indicates that the DBSC also serves as a useful and effective result control through the use of pay for performance and perceptions of fairness and legitimacy that create motivation and support conformity. Measurability of performance and setting

performance targets can be helpful to establishing a climate of result control, but “softer” considerations such as the perceived fairness and legitimacy of the PMM also appear to be important to its effectiveness as a result control and pay-for-performance system. The perceived relational properties found here, combined with other attributes of this PMM and acceptable feedback from financial success, appear to be sufficient to support continued PMM use.

Limitations

Our study has failed to support the normative assumption of cause-and-effect relations in a PMM at a business-unit level. At a minimum, our study has refuted cause and effect as an explanation for the continued use of this company’s DBSC (Popper 1959, 1963). As in all case research, one can question the reproducibility of the results, but statistical support for cause and effect will be elusive in the best of circumstances because of incompleteness of PMM, managers’ adaptations to feedback, and instability of firms’ production functions.

This study is limited by the quarterly data that might disguise shorter response times among leading and lagging performance measures. Some measures once thought to be important to the performance model were dropped by the company for measurement deficiency reasons. These omissions might cause material bias in estimated statistical relations, if in fact they are important to explaining overall performance. The data are limited to the distribution channel of the company’s value chain, but overall profit accrues to the entire chain. Thus, distributor profitability, which is tightly controlled by the company, might not reflect the full distribution contributions to overall profitability.

Future Research

We suspect that archival PMM data from most organizations will be similarly messy for several reasons. First, thorough research and development of PMM measures might be impractical given the strategic urgency of implementing a new PMM. Learning by doing and continual improvement seem likely. Second, strategic and operational changes will occasion changes in the PMM. Third, one should expect firms to take actions based on PMM results to improve the organization, which will change the data-generating processes. Because all of these changes to the production function and interruptions to the time-series of data are likely in dynamic organizations, one should not expect anything like laboratory conditions and measurements. If a firm intends and even achieves a causal PMM in the real world of dynamic organizations and periodic data collection, cause and effect might not be observable or testable. Thus, tests for cause and effect may not be useful for judging even intentionally causal PMM.

We acknowledge that challenging earlier results with critical argumentation and other types of data are important for advancing our knowledge of these phenomena; however a study’s methodology must fit

the nature of the problem (Popper 1961, 1963; Norreklit et al., forthcoming). If the logic of financial accounting forms a crucial part of a PMM, one must design an empirical study to reflect the business logic of the company. Although logical analysis can refute logical arguments, we caution that while qualitative analysis is suggestive it may be insufficient by itself to support or refute empirical hypotheses. Thus, we believe that dialogue-based research methods complement statistical tests of cause-and-effect.

We agree with Ittner and Larcker (2003) that firms and researchers should examine PMM relations between means and ends and carefully estimate the financial consequences of alternative actions. Only the rare firm living in a stable environment may be able to establish a predictable, cause-and-effect business model. Because for most firms the business context is dynamic and does not follow mechanical laws, firms may intentionally, but perhaps without regard to labels and their implications for validation, create PMM that cannot be validated statistically. Thus, estimating effects and predicting future performance of logical and finality relations or changing cause-and-effect relations must depend on more than extrapolations of prior results. Not only past results but also the financial impacts of future opportunities should form part of performance prediction, and inevitably management must make subjective assumptions and judgments. Evaluating the validity of PMM may require logical, qualitative and financial cost-benefit analyses (including business-model simulations); the statistical tools of normal science may not apply easily.

On a practical level, more work might be justified to improve existing PMM measures and accuracy of reporting and to reconfigure PMM as the organization gains experience and expertise. Consistent commitment and fine-tuning might improve its statistical reliability and predictive ability over time (e.g., Shields and Young, 1989), particularly in PMM that reflect physical processes and possibly for finality relations such as those involving customer satisfaction. However, if logical and finality relations are relatively frequent, financial, cost-benefit analysis will be more important to judging the reliability of PMM than statistical analysis. It is possible that companies care more that the PMM tells an intuitive story and provides an accepted and effective basis for result control than whether the PMM embodies statistically significant relations throughout. In the case studied here, for example, the firm might focus on establishing the fairness and legitimacy of the DBSC in its foreign distributorships before deploying it globally. The firm also could investigate the financial cost-benefit behind the consistent logical result that distributor profitability lags sales growth by a full year. Perhaps seasonality drives this lag, but perhaps the company and its distributors could learn how to make their new customers profitable more quickly.

Based on the summary of relations shown in Figure 4, we pose the following “climate of control” propositions for consideration by future research:

P1: An organization’s climate of control influences the **design** of PMM.

- Factors include management style (e.g., pay for performance), strategic goals, and the use of accounting tools.
- Performance measure relations in PMM are functions of contingencies such as desired climate of control and environmental uncertainty.
- Climate of control and beliefs about relations among performance measures interact to affect the design of PMM.

P2: The design of a PMM affects its **use**

- Business model communication is moderated by the types of relations imbedded in the PMM.
- Business model communication generates control legitimacy, fairness, and learning that affect motivation, conformity and goal congruence within the organization, moderated by the business model's predictive ability.
- Business model predictive ability is moderated by the types of relations imbedded in the PMM.
- Business model predictive ability and goal congruence affect decision effectiveness.

P3: PMM design and use are influenced by **financial feedback** because all elements of the proposed climate of control theory of are **dynamic**.

Although PMM such as the BSC have spanned the globe and appear in every type of business, government, and nongovernmental organization, we have much to learn about how complex PMM are used. Future research also can investigate conditions where logical and finality relations are expected to complement or supplant cause-and-effect relations, or vice-versa. We have witnessed what appears to be substitution of finality and logical relations for cause-and-effect relations in a predominantly service-oriented PMM. Whether this is intentional or common is unknown to us, but we suspect that many, perhaps most PMM will tend to have few unambiguous cause-and-effect relations. Cause-and-effect relations might be common in the PMM of organizations that are strongly based on physical processes, such as those in extractive and manufacturing industries. Service-oriented organizations or those parts of large organizations that are largely service, it appears to us, may be far more likely to construct PMM with finality relations. Logical relations that link upstream outcomes to financial outcomes may be equally likely in all types of organizations. Given that companies operate in a context of accounting performance, we know that the measurements have to be linked to financial performance one way or another, but we do not know much about how the links are constructed and made operational. We also do not know whether PMM success and ultimately organizational success are positively associated with complementary use of all types of relations or whether focus on one or another increases PMM success. We look forward to future research to further examine these issues to better our understanding of this complex phenomenon.

REFERENCES

- Abernethy, M., A. Lillis. 2001. Interdependencies in organization design: A test in hospitals. *Journal of Management Accounting Research* 13: 107-129.
- Abernethy, M., M. Horne, A. Lillis, M. Malina, and F. Selto. 2005. "A multi-method approach to building causal performance maps from expert knowledge." *Management Accounting Research* 16: 135-155
- Alvarez, J.E. 1998. *The Diffusion and Consumption of Business Knowledge*, London, Macmillan Press.
- Ambrosini, V. and C. Bowman. 2002. Mapping successful organizational routines. in Huff, A. and M. Jenkins (eds.). (2002). *Mapping Strategic Knowledge*. Thousand Oaks, CA: Sage Publications: 19-45.
- Anthony, R. and V. Govindarajan. 1998. *Management Control*. Burr Ridge, IL: Irwin.
- Arbnor, I. and Bjerke, B., 1997. *Methodology for Creating Business Knowledge*, London: Sage Publications.
- Ashley, R. C.W.J. Granger, and R. Schmalensee. 1980. Advertising and aggregate consumption: An analysis of causality. *Econometrica* 48: 1149-67.
- Baker, G. 2002. Distortion and risk in optimal incentive contracts. *Journal of Human Resources*, 37 (4): 728-751.
- Banker, R., G. Potter, D. Srinivasan. 2000. An empirical investigation of an incentive plan that includes nonfinancial performance measures. *The Accounting Review*. 75(1): 65 -92.
- Bollen, K. 1989. *Structural Equations with Latent Variables*. New York: Wiley.
- Bourguignon, A., V. Malleret, and H. Nørreklit. 2004. The American balanced scorecard versus the French tableau de bord: The ideological dimension. *Management Accounting Research*. 15, (2): 107
- Bryant, L., D. Jones, S. Widener. 2004. Managing value creation within the firm: An examination of multiple performance measures. *Journal of Management Accounting Research*. 16: 107-31
- Burney, L., C. Henle, and S. Widener. 2006. Do characteristics of strategic performance measurement systems used in incentives enhance organizational fairness? Rice University working paper.
- Carruthers, B. G. 1995. Accounting, ambiguity, and the new institutionalism. *Accounting, Organizations and Society*, 20: 313-328.
- Chenhall, R. 2003. Management control systems design within its organizational context: findings from contingency-based research and directions for the future. *Accounting, Organizations and Society* 28(2-3): 127-168.
- Corbin, J., A. Strauss. 1990 Grounded Theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13 (1): 3-21.

- Covaleski, M. A., Dirsmith, M.W. and Samuel, S. 1996. Managerial accounting research: The contributions of organizational and sociological theories. *Journal of Management Accounting Research*, 8:1-35.
- Cook, T. and D. Campbell. 1979. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*.
- Darnell, A. 1994. *A Dictionary of Econometrics*. Hants, England: Edward Elgar Publishing Limited.
- Datar, S., S. Culp and R. Lambert. 2001. Balancing performance measures. *Journal of Accounting Research* 39(1): 75 – 94.
- Dearden, J. 1969. The case against ROI control. *Harvard Business Review* 47: 124-35.
- de Geus, A. 1994. Modeling to predict or to learn? In Morecroft and Sterman (eds.) *Modeling for Learning Organizations*. Portland: Productivity Press.
- Drazin, R. and A. van de Ven. 1985. Alternative forms of fit in contingency theory. *Administrative Science Quarterly*, V. 30 Iss. 4: 514-39.
- Dougherty, D. 2002. Grounded theory research methods. *Blackwell Companion to Organizations*: 849-866.
- Eccles, R. 1991. The performance measurement manifesto. *Harvard Business Review* 69: 131-7
- Edwards, P. 1972. *The Encyclopaedia of Philosophy*, Macmillan Publishing Co., Inc. & The Free Press, US, Vol. 1-8.
- Feltham, G.A. and J. Xie. 1994. Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review*. Sarasota 69(3): 429-55.
- Forrester, J. 1994. Policies, decisions, and information sources for modeling. In Morecroft and Sterman (eds.) *Modeling for Learning Organizations*. Portland: Productivity Press.
- Foucault M. 2000. “Governmentality” and “The subject and power”, in Faubion, J.D. (ed) *Michel Foucault: Power – The Essential Works*, vol. three, New York: The New Press: 201-222, 326-348.
- Frigo, M. 2002a. Nonfinancial performance measures and strategy execution. *Strategic Finance* Aug: 6-9.
- Frigo, M. 2002b. Strategy-focused performance measures. *Strategic Finance* Sep: 10, 14-15.
- Granger, C.W.J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37(3): 424-438.
- Granger, C.W.J. 1980. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control* 2(4): 329-352.
- Green, T. 1992. *Performance and Motivation Strategies for Today's Workforce: A Guide to Expectancy Theory Applications*. Westport, CT: Greenwood Publishing Group.
- Huff, A. and M. Jenkins (eds.). 2002. *Mapping Strategic Knowledge*. London: Sage Publications.
- Ijiri, Y. 1978. *The Foundations of Accounting Measurement: A Mathematical, Economic, and Behavioral Inquiry*. Houston, TX: Scholars Book Co.

- Iribarne (d'), P. 1994. The honour principle in the 'Bureaucratic Phenomenon'. *Organization Studies* 15/1, 81-97.
- Ittner, C. and D. Larcker. 1998. Are non-financial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research* 36 (supplement): 1-35.
- Ittner, C. and D. Larcker. 2001. Assessing empirical research in managerial accounting: A value-based management perspective. *Journal of Accounting and Economics*. 32(1-3): 349-410.
- Ittner, C., and D. Larcker. 2003. Coming up short on nonfinancial performance measurement. *Harvard Business Review* v. 81, n.11: 88-95.
- Ittner, C., D. Larcker and M. Meyer. 2003. Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review* 78(3): 725-58.
- Ittner, C., D. Larcker and T. Randall. 2003b. Performance implications of strategic performance measurement in financial services firms. *Accounting, Organizations and Society*. 28(7, 8): 715-41.
- Johnston, J. 1994. *Econometric Methods* 3rd edition. New York: McGraw Hill Book Company.
- Kaplan, R.S. and R. Cooper. 1998. *Cost & Effect: Using Integrated Costs to Drive Profitability and Performance*, Harvard Business School Press, Boston.
- Kaplan, R. and D. Norton. 1992 The balanced scorecard – measures that drive performance. *Harvard Business Review* January – February: 71-79.
- Kaplan, R. and D. Norton. 1996. *The Balanced Scorecard*. Boston, MA: Harvard Business School Press.
- Kaplan, R. and D. Norton. 2001. *The Strategy-Focused Organization*. Boston, MA: Harvard Business School Press.
- Lipe, M. and S. Salterio. 2002. A note on the judgmental effects of the balanced scorecard's information organization. *Accounting, Organizations & Society*, 27(6): 531-40.
- Locke E. and G. Latham. 1990. *A Theory of Goal Setting & Task Performance*. Englewood Cliffs, N.J.: Prentice Hall
- Luft, J. and M. Shields. 2002. Learning the drivers of financial performance: Judgment and decision effects of financial measures, nonfinancial measures, and statistical models. Michigan State University working paper.
- Magretta, J. 2002. Why business models matter. *Harvard Business Review*, 80(5).
- Malina, M. and F. Selto. 2001. Communicating and controlling strategy: An empirical study of the effectiveness of the balanced scorecard. *Journal of Management Accounting Research* 13: 48 - 90.
- Malina, M. and F. Selto. 2004. Choice and change of performance model measures. *Management Accounting Research*. 15(4): 441-60.
- Mattessich, R. 1995. Conditional-normative accounting methodology: Incorporating value judgments and means-end relations of applied science. *Accounting, Organizations and Society* 20: 259-285.

- Merchant, K. 1998. *Modern Management Control Systems*. Upper Saddle River, NJ: Prentice Hall.
- Meyer J. and Rowan, B. 1977. Institutionalized organizations: Formal structure of myth and ceremony, *American Journal of Sociology*, 80, 340-363.
- Miller P. & T. O'Leary, 1987. "Accounting and the construction of the governable person," *Accounting, Organizations and Society*, 12 (3), 235-265.
- Morecroft, J. and J. Sterman, eds. 1994. *Modeling for Learning Organizations*. Portland, OR: Productivity Press.
- Morecroft, J., R. Sanchez, and A. Heene (eds). 2002. *Systems Perspectives on Resources, Capabilities, and Management Processes*. Amsterdam: Pergamon.
- Nonaka, I. 1994. A dynamic theory of organizational knowledge creation. *Organization Science*, 5(1): 14-38.
- Nonaka, I. and H. Takeuchi. 1995. *The Knowledge-Creating Company*. New York: Oxford University Press.
- Nørreklit, H. 2000. The balance on the balanced scorecard: A critical analysis of some of its assumptions. *Management Accounting Research* 11: 65-88.
- Nørreklit, L. 1987. *Formal Structures in Social Logic*. Aalborg, DK: Aalborg University Press.
- Nørreklit, L., Nørreklit, H., and P. Israelsen. 2006. Validity of management control topoi: Towards constructivist pragmatism. *Management Accounting Research*. 17(1): 42-71.
- Nørreklit, H., Nørreklit, L., and F. Mitchell. forthcoming. Theoretical conditions for validity in accounting performance measurement, in Neely, A. (ed.), *Business Performance Measurement - Frameworks and Methodologies*, Cambridge University Press,
- O'Connor, G., M. Rice, L. Peters, R. Veryzer. 2003. Managing interdisciplinary, longitudinal research teams: Extending grounded theory-building methodologies. *Organization Science* 14(4): 353-373.
- Popper K. 1959. *The Logic of Scientific Discovery*. (translation of *Logik der Forschung*). Hutchinson, London.
- Popper K. 1961. *The Poverty of Historicism* (2nd. ed). London: Routledge.
- Popper K. 1963, *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge, London.
- Porter, M. 1985. *Competitive Advantage*. New York: The Free Press.
- Porter, T. M. 1995. *Trust in Numbers*. Princeton, NJ: Princeton University Press.
- Ridgway, V. F. 1956, Dysfunctional consequences of performance measurements. *Administrative Science Quarterly* 1: 240-247.
- Rucci, A., S. Kirn and R. Quinn. 1998. The employee-customer-profit chain at Sears. *Harvard Business Review* 76(1): 82-97.
- Sanchez, R., A. Heene and H. Thomas (eds.). 1996. *Dynamics of Competence-Based Competition*. Oxford: Pergamon.

- Searle, J. R. 1995. *The Construction of Social Reality*. New York: Free Press.
- Shields, M. and S.M. Young. 1989. A behavioral model for implementing cost management systems. *Journal of Cost Management* (2): 29-34.
- Simons, R. 2000. *Performance Measurement & Control Systems for Implementing Strategy*. Upper Saddle River, NJ: Prentice Hall.
- Slife, B.D. and R.N. Williams. 1995. *What's behind the research? Discovering hidden assumptions in the behavioral sciences*. Sage Publications, London.
- Willard, B. 2005. *The NEXT sustainability wave: Building boardroom buy-in*. British Columbia, Canada: New Society Publishers.
- Zimmerman, J. 1997. *Accounting for Decision Making and Control*. Burr Ridge, IL: Irwin-McGraw-Hill.

FIGURE 1
DBSC Timeline

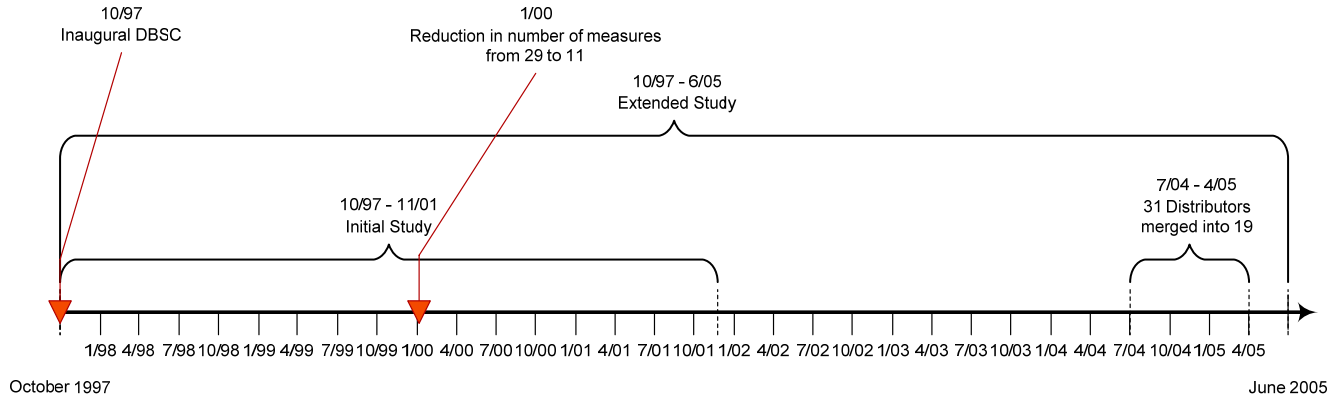
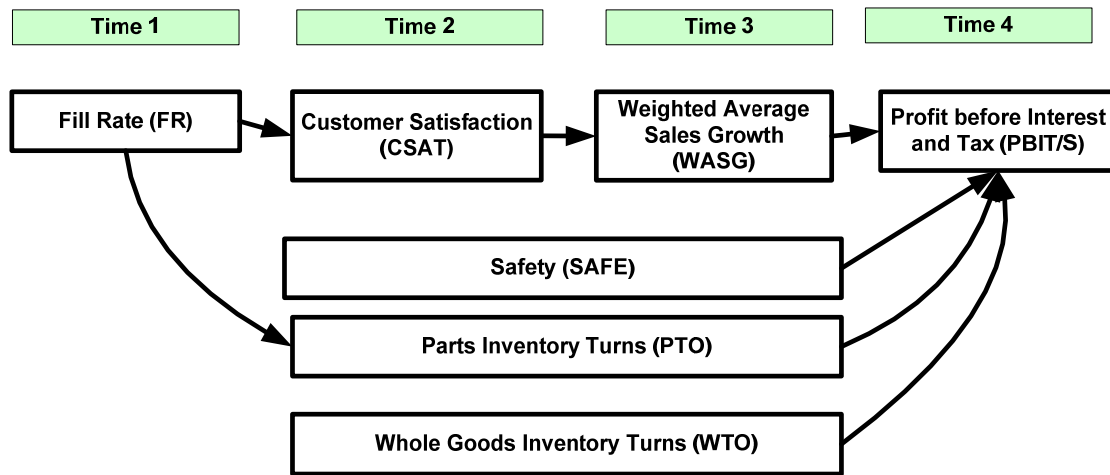


FIGURE 2

Management's and Distributors' Expected DBSC Relations



Narrative Summary of Figure 2: The distributor's customer fill rate affects parts inventory turnover. Order fill rate is expected to affect customer satisfaction, because parts availability affects how quickly the distributor can meet customers' parts and service needs. Note that the company's measure of customer satisfaction is obtained during the quarter it is reported, and it might have more immediate impact than is observable, just by construction. The company believes the best means to drive sales growth (and distributor profitability) is through improved customer satisfaction. Safety affects profitability through insurance costs and lost billable time. Safety and the turnover of inventories have direct impacts on distributor profitability. Note: Time periods are relative, and are not intended to accurately reflect quarterly effects.

Source: Coded interview transcripts from Malina and Selto (2001).

FIGURE 3

Time Series of Customer Satisfaction Performance, Q1 – Q31

Figure 3A: Customer Satisfaction, Percent Red/Yellow/Green, All Distributors

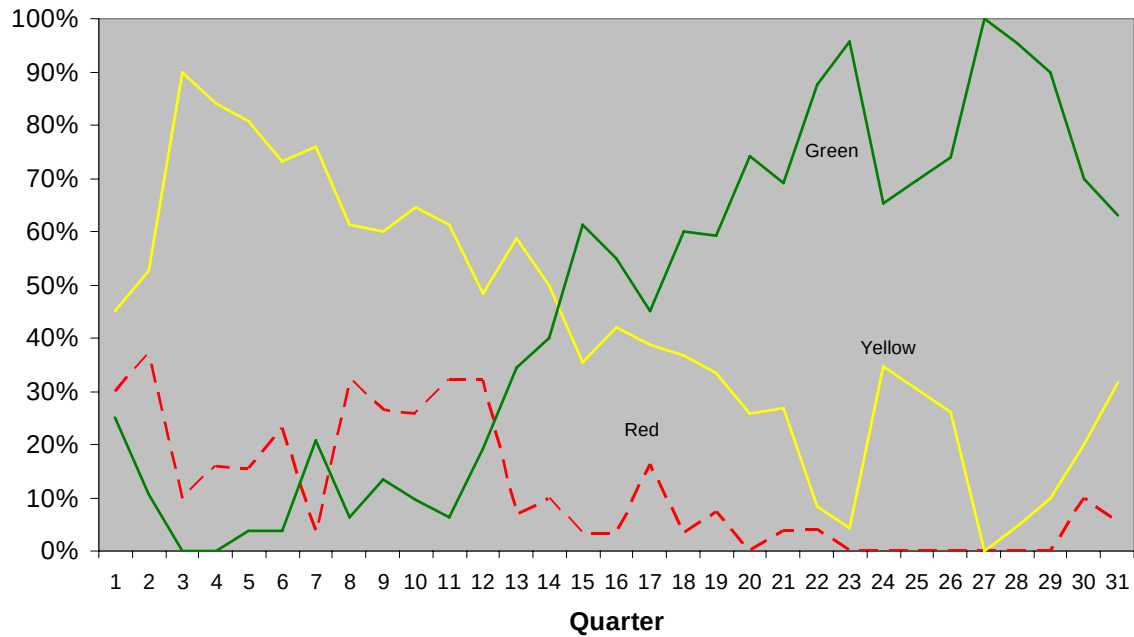


Figure 3B: Error-Bar Chart of Customer Satisfaction Performance, All Distributors

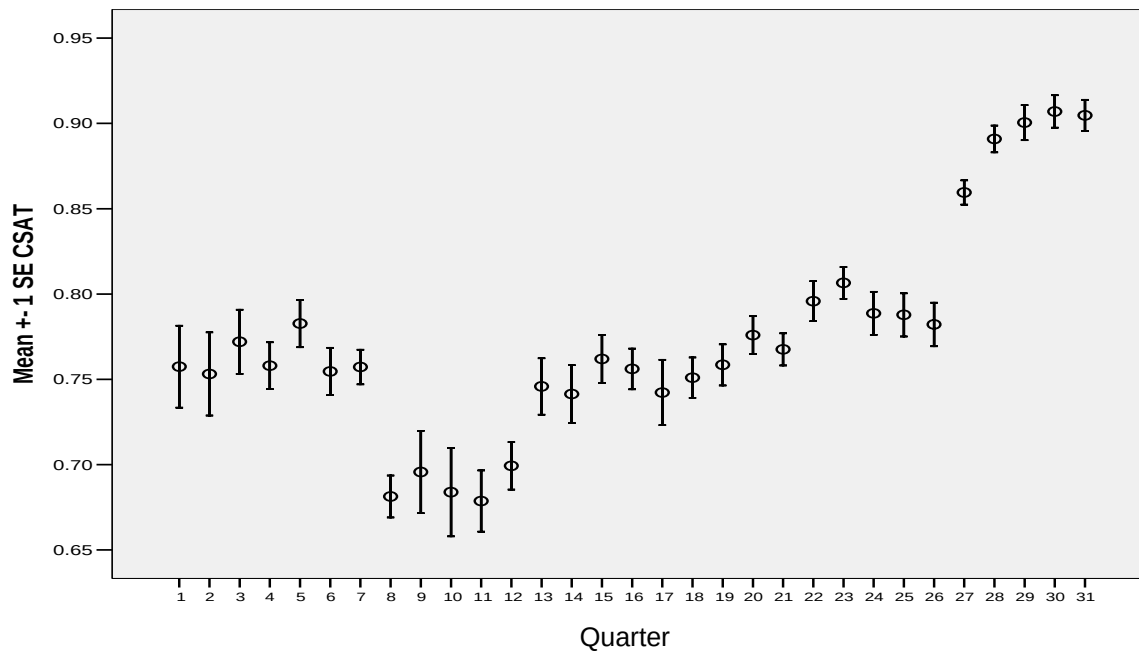
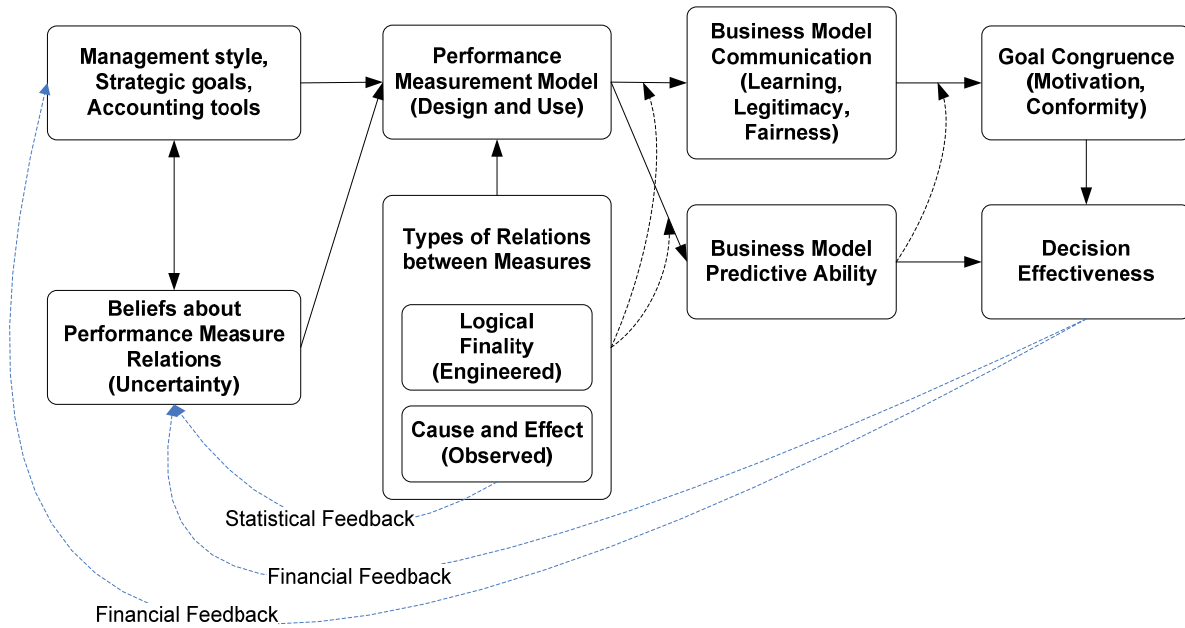


FIGURE 4
Model of PMM Control Effectiveness Theory



Note that intermediate feedback loops also can and probably do exist.

TABLE 1**Distributor Balanced Scorecard Quarterly Measures**

DBSC Measure (testable measure abbreviation)	Definition	Source at the time of the Research
Customer satisfaction (CSAT).....	Score on customer satisfaction event card	External, third party survey
First-time customer fill rate (FR).....	Percentage of parts ordered by customers filled within 24 hours	Distributor information system
Parts inventory turns (PTO).....	Parts cost of sales divided by average parts inventory cost	Company information system
Distributor profitability (PBIT/S).....	Distributor profit before interest and taxes, as a percentage of sales	Company information system
Safety (SAFE).....	Lost-time accidents per 200,000 hours worked	Distributor information system
Weighted average sales growth (WASG)...	Created from factor scores of three sales growth figures (parts, service, and other)	Company information system
Whole goods inventory turns (WTO).....	Whole goods cost of sales divided by average whole goods inventory cost	Company information system

Note that the variable weighted average sales growth, WASG, was created for this research from factor analysis of the three available sales growth measures (parts, service, and other) that yielded a single common factor with strong loadings by the three components. All statistical analyses were repeated with disaggregated sales growth figures with no materially or statistically different results than those reported here. The lack of difference in results by type of sales growth indicates no significant effects from product mix differences across distributors.

TABLE 2

Initial Qualitative Analysis of Comments Referring to Relations among DBSC Measures

	Company Managers					Distributors									
CODES	J	K	L	M	N	A	B	C	D	E	F	G	H	I	Totals
Panel A: Relations Cause & effect	9	9	10	11	7	18	10	19	15	11	26	13	12	9	179
Panel B: Measures															
Customer satisfaction (CSAT)	3	0	1	2	1	5	1	5	6	2	8	5	9	3	51
Parts fill rate (FR)	1	1	0	0	0	4	1	3	1	1	4	2	1	3	22
Weighted average sales growth (WASG)	3	5	5	11	0	5	3	6	4	3	4	3	4	3	59
Parts or whole goods inventory turnover (PTO, WTO)	2	0	0	0	0	0	0	0	2	2	2	0	2	0	10
Profit before interest and taxes divided by sales (PBIT/S)	2	1	4	4	1	0	4	3	6	5	2	1	4	0	37
Safety (SAFE)	1	0	0	0	15	1	2	1	1	2	1	1	1	2	28

Panel A: Each number refers to the frequency with which respondent referred to cause-and-effect relations, as initially judged by the authors. Eighty-four of these referred to relations between two specific measures and were used to construct the DBSC model and system of path equations.

Panel B: Each number refers to the frequency with which each respondent referred to this measure.

TABLE 3
Granger Estimations for DBSC Equations (Quarters 1 through 14)

Panel A: Dependent variable = PTO (Equation 1)

(1)	(2)		(3)		(4)		(5)		(6)	
	Granger		Granger w/fixed		Log-linear		1 Qtr Change		4 Qtr Change	
	B	t-stat	B	t-stat	B	t-stat	B	t-stat	B	t-stat
(Constant)	.164	.423	.714	.930	.082	1.623	-.060	-2.018*	.229	3.247**
PTO1	.927	10.869***	.512	5.884***	.940	11.065***				
PTO2	.078	.654	.061	.601	.102	.838				
PTO3	-.121	-1.099	-.071	-.758	-.193	-1.653				
PTO4	.045	.597	-.146	-.777	.093	1.160				
FR	.665	1.325	.887	1.403	.112	1.144	.032	.083	.415	.642
FR2	-.032	-.067	-.093	-.204	.006	.049				
FR4	-.514	-1.207	.130	.324	-.057	-.682				
Fixed effects			23 sig. fixed effects							
Adjusted R ²	.835		.882		.870		.004		.003	

NB: FR, FR1, FR3 are highly collinear ($R > .70$)

Panel B: Dependent variable = CSAT (Equation 2)

(1)	(2)		(3)		(4)		(5)		(6)	
	Granger		Granger w/fixed		Log-linear		1 Qtr Change		4 Qtr Change	
	B	t-stat	B	t-stat	B	t-stat	B	t-stat	B	t-stat
(Constant)	.140	1.541	.575	3.503**	-.074	-2.000*	.006	1.096	-.014	-1.349
CSAT1	.507	5.660***	.010	.099	.488	5.518***				
CSAT2	.180	1.697	-.150	-1.339	.176	1.756				
CSAT3	.030	.282	-.140	-1.325	.053	.513				
CSAT4	-.023	.237	-.091	-.872	-.043	-.480				
FR	.189	2.001*	.414	3.098**	.191	1.919	.188	2.731**	.158	1.549
FR2	-.122	-1.355	.031	.335	-.110	-1.132				
FR4	.041	.520	-.064	-.794	.030	.364				
Fixed effects			15 sig. fixed effects							
Adjusted R ²	.359		.489		.354		.026		.009	

NB: FR, FR1, FR3 are highly collinear ($R > .70$)

Model	Model description
Granger	Granger estimation models reported in the paper, up to 14 observations per distributor
Granger w/fixed..	Granger models w/30 fixed distributor effects (not shown), up to 14 observations per distributor
Log-linear.....	Log transformed models, no fixed effects, up to 14 observations per distributor
1 Qtr Change.....	Changes model, first differences, up to 13 observations per distributor
4 Qtr Change.....	Changes model, fourth differences, up to 10 observations per distributor

Variables in shaded area are hypothesized causes of performance. Coefficients in **bold** font are significant and signed as predicted (*** < .001, ** < .01, * < .05)

NB: Only significant, lagged observations of performance drivers might be interpreted as causally related to performance (shaded rows). Because first or fourth-difference variables also contain the contemporaneous value, their coefficients are ambiguous about causality.

Table 3 (continued)

Panel C: Dependent variable = WASG (Equation 3)

(1)	(2)		(3)		(4)		(5)		(6)	
	Granger		Granger w/fixed		Log-linear		1 Qtr Change		4 Qtr Change	
	B	t-stat	B	t-stat	B	t-stat	B	t-stat	B	t-stat
(Constant)	.057	.399	.014	.049	-.118	-.193	-.124	-1.294	.101	1.847
WASG1	.650	8.621***	.488	5.896***	1.452	5.946***				
WASG2	-.200	-3.539**	-.122	-2.084*	-.169	-.597				
WASG3	.051	1.255	.020	4.91***	.437	1.584				
WASG4	.010	.472	.024	1.069	-.647	-2.945**				
CSAT	-.046	-.288	.042	.181	-1.791	-1.423	.099	.086	.123	.259
CSAT2	.041	.200	-.002	-.007	1.275	.835				
CSAT3	.113	.602	.114	.533	-.532	-.344				
CSAT4	-.049	-.271	-.045	-.335	1.893	1.369				
Fixed effects			1 sig. fixed effect							
Adjusted R ²		.392		.409	.302		.003		.004	

NB: CSAT and CSAT1 are highly collinear ($R > .70$)

Panel D: Dependent variable = PBIT/S (Equation 4)

(1)	(2)		(3)		(4)		(5)		(6)	
	Granger		Granger w/fixed		Log-linear		1 Qtr Change		4 Qtr Change	
	B	t-stat	B	t-stat	B	t-stat	B	t-stat	B	t-stat
(Constant)	.012	1.520	.067	2.371*	-.251	-.400	.002	1.379	-.004	-,2.306*
PBIT1	.339	3.366**	.166	1.614	.376	2.628*				
PBIT2	.075	.539	-.134	-.984	.093	.709				
PBIT3	.267	2.242*	.011	.088	.086	.987				
PBIT4	.195	1.810	.070	.560	.187	2.264*				
WASG	.035	4.304***	.032	3.189**	.058	2.869**	.002	.825	.005	.889
WASG1	.001	.055	-.066	-.664	-.011	-.132				
WASG2	-.007	-.768	-.002	-.254	.041	.456				
WASG3	-.015	-1.688	-.005	-.647	-.048	-.556				
WASG4	.043	5.135***	.056	6.176***	.149	1.910				
PTO	-.004	-1.826	-.001	-.265	-.180	-.866	.005	1.760	-.001	-.697
PTO4	.001	.458	-.001	-.345	-.024	-.111				
WTO	.0004	.071	.001	1.054	.243	1.861	.002	2.071*	.001	1.706
WTO4	-.001	-1.274	-.001	-1.174	-.281	-2.020*				
SAFETY	.000	.622	.000	.491	.007	.408	-.000	-.697	.001	1.365
SAFE1	.000	.537	.00007	.081	.086	.895				
SAFE2	.000	.410	-.00005	-.002	.036	.409				
SAFE3	-.001	-1.173	-.001	-.984	.012	.135				
SAFE4	-.001	-1.382	-.001	-1.009	-.038	-.613				
Fixed effects			5 sig. fixed effect							
Adjusted R ²		.392		.409	.302		.003		.004	

NB: PTO1, 2, 3, and 4 and WTO1, 2, 3, and 4 are highly collinear ($R > .70$)

TABLE 4
Performance Measure Descriptive Statistics
Full Dataset (Q1 – Q31)

Performance Measures	N	Mean	Min	Max	Std Dev
FR	760	.820	.010	.978	.075
PTO	856	4.462	1.100	25.300	1.575
WTO	856	8.998	1.300	36.400	4.647
SAFE	736	3.202	.000	23.100	2.815
CSAT	784	.766	.000	1.000	.096
WASG*	856	.235	-.533	27.390	1.053
PBIT	855	.048	-.016	.206	.023
Complete N (listwise)	700				

* Includes five outlying observations of WASG

CSAT – Customer satisfaction

FR – Parts fill rate

PTO – Parts inventory turns

PBIT/S – Distributor profit before income tax as a percent of sales

SAFE – Safety

WTO – Whole goods inventory turns

WASG – Weighted average sales growth

TABLE 5
Pairwise Pearson Correlations of Unlagged Variables
Full Dataset (Q1 – Q31)

	(Q1 – Q31; $707 \leq N \leq 856$)					
	PTO	WTO	SAFE	CSAT	WASG	PBIT/S
FR	.093*	-.072*	.094*	.088*	-.081*	.107**
PTO		.363**	.046	.150**	-.019	.221**
WTO			-.030	.012	.001	.178**
SAFE				-.008	.163**	.115**
CSAT					.040	.038
WASG						-.015

* Correlation is significant at $\alpha = 0.05$ (2-tailed).

** Correlation is significant at $\alpha = 0.01$ (2-tailed)

CSAT – Customer satisfaction

FR – Parts fill rate

PTO – Parts inventory turns

PBIT/S – Profit before income tax as a percent of sales

SAFE – Safety

WTO – Whole goods inventory turns

WASG – Weighted average sales growth

TABLE 6

Granger Estimations for DBSC Equations (Quarters 1 through 28)

Panel A: Dependent variable = PTO (Equation 1)

(1)	(2) Granger		(3) Granger w/fixed		(4) Log-linear		(5) 1 Qtr Change		(6) 4 Qtr Change	
	B	t-stat	B	t-stat	B	t-stat	B	t-stat	B	t-stat
(Constant)	0.413	.746	0.560	0.647	0.086	2.723**	0.87	2.803*	-0.350	-6.677***
PTO1	1.001	14.446***	0.918	13.996***	0.927	21.154***				
PTO2	-.050	-.537	-.018	-.211	0.086	1.451				
PTO3	0.306	3.168**	0.223	2.533**	-.052	-.877				
PTO4	-.252	-3.294**	-.192	-2.740**	-.002	-.036				
FR	-.424	-.523	-.445	0.607	0.038	0.439	-.134	-.817	1.611	2.222*
FR2	0.865	1.063	0.942	1.117	0.128	1.490				
FR4	-.846	-1.370	-.875	-1.350	-.141	-2.162*				
Fixed effects			1 sig. fixed effect							
Adjusted R ²	0.742		0.759		0.878		0.001		0.007	

NB: FR, FR1, FR3 are highly collinear ($R > .70$)

Panel B: Dependent variable = CSAT (Equation 2)

(1)	(2) Granger		(3) Granger w/fixed		(4) Log-linear		(5) 1 Qtr Change		(6) 4 Qtr Change	
	B	t-stat	B	t-stat	B	t-stat	B	t-stat	B	t-stat
(Constant)	0.109	2.218*	0.053	0.819	-.012	-.889	0.007	2.62*	-.033	-8.22***
CSAT1	0.508	10.952***	0.533	12.194***	0.532	12.629***				
CSAT2	0.161	3.194**	0.185	3.831***	0.200	4.348***				
CSAT3	0.038	0.752	0.022	0.462	0.059	1.316				
CSAT4	0.069	1.534	0.031	0.705	0.046	1.132				
FR	0.203	3.176**	0.194	2.932**	0.184	2.950**	0.180	3.381*	0.256	4.165***
FR2	-.112	-1.718	-.081	-1.257	-.101	-1.588				
FR4	-.003	-.063	0.013	0.255	-.006	-.114				
Fixed effects			3 sig. fixed effects							
Adjusted R ²	0.433		0.553		0.544		0.016			

NB: FR, FR1, FR3 are highly collinear ($R > .70$)

Model	Model description
Granger	Granger estimation models reported in the paper, up to 14 observations per distributor
Granger w/fixed..	Granger models w/30 fixed distributor effects (not shown), up to 14 observations per distributor
Log-linear.....	Log transformed models, no fixed effects, up to 14 observations per distributor
1 Qtr Change.....	Changes model, first differences, up to 13 observations per distributor
4 Qtr Change.....	Changes model, fourth differences, up to 10 observations per distributor

Variables in shaded area are hypothesized causes of performance. Coefficients in **bold** font are significant and signed as predicted (*** < .001, ** < .01, * < .05)

NB: Only significant, lagged observations of performance drivers might be interpreted as causally related to performance (shaded rows). Because first or fourth-difference variables also contain the contemporaneous value, their coefficients are ambiguous about causality.

Table 6 (continued)

Panel C: Dependent variable = WASG (Equation 3)

(1)	(2)		(3)		(4)		(5)		(6)	
	Granger		Granger w/fixed		Log-linear		1 Qtr Change		4 Qtr Change	
	B	t-stat	B	t-stat	B	t-stat	B	t-stat	B	t-stat
(Constant)	0.053	0.680	-0.042	0.597	-0.195	0.481	-0.099	-1.323	0.051	2.363*
WASG1	0.746	18.057***	0.692	17.238***	1.230	10.248***				
WASG2	-0.208	-5.605***	-0.185	-5.168***	0.046	0.284				
WASG3	0.044	1.451	0.041	1.377	0.078	0.473				
WASG4	0.004	0.223	0.004	0.259	-0.325	-2.626*				
CSAT	-0.057	-0.643	-0.014	-0.170	-1.435	-1.400	-0.015	-0.017	0.198	0.953
CSAT2	-0.022	-0.198	-0.017	-0.165	0.709	0.578				
CSAT3	0.035	0.327	0.055	0.584	-0.887	-0.716				
CSAT4	0.052	0.519	0.105	0.273	3.001	2.646**				
Fixed effects			0 sig. fixed effects							
Adjusted R ²	0.487		0.480		0.327		0.002		0.000	

NB: CSAT and CSAT1 are highly collinear ($R > .70$)

Panel D: Dependent variable = PBIT/S (Equation 4)

(1)	(2)		(3)		(4)		(5)		(6)	
	Granger		Granger w/fixed		Log-linear		1 Qtr Change		4 Qtr Change	
	B	t-stat	B	t-stat	B	t-stat	B	t-stat	B	t-stat
(Constant)	-0.004	-1.214	0.200	3.268**	-0.321	-0.775	0.001	.822	-0.002	-1.642
PBIT1	0.552	11.893**	0.388	8.770***	0.726	6.079***				
PBIT2	0.039	0.741	-0.018	-3.73	-0.067	-0.652				
PBIT3	0.081	1.563	0.022	0.632	0.169	2.019*				
PBIT4	0.171	3.700**	0.086	2.026*	0.047	0.593				
WASG	0.034	6.601***	0.034	7.185***	0.012	0.667	0.003	1.444	0.012	3.324**
WASG1	-0.008	-1.344	-0.007	-1.441	0.112	2.378*				
WASG2	-0.010	-1.785	-0.007	-1.472	-0.001	-0.013				
WASG3	-0.003	-0.565	-0.001	-0.205	-0.029	-0.501				
WASG4	0.020	4.153***	0.027	6.199***	0.069	1.529				
PTO	0.002	2.819**	0.002	3.578***	-0.133	-0.786	0.004	1.488	0.000	0.585
PTO4	0.000	0.238	0.001	0.903	0.306	1.713				
WTO	0.000	0.206	0.000	0.881	0.128	1.263	0.001	2.745*	0.000	0.898
WTO4	-0.000	-0.002	0.000	1.066	-0.130	-1.158				
SAFETY	0.001	1.522	0.000	0.560	0.012	0.682	-0.000	-0.530	0.001	3.322**
SAFE1	0.000	-0.914	-0.000	-0.902	-0.055	-0.681				
SAFE2	0.000	0.424	0.000	0.017	0.065	0.784				
SAFE3	-0.000	-0.200	-0.000	-0.051	-0.025	-0.317				
SAFE4	0.000	-1.361	-0.001	-2.555*	-0.029	-0.467				
Fixed effects			21 sig. fixed effects							
Adjusted R ²	0.580		0.646		0.375		0.024		0.037	

NB: PTO1, 2, 3, and 4 and WTO1, 2, 3, and 4 are highly collinear ($R > .70$)

TABLE 7
Granger Predictive Ability Results

Dependent Variable	Root Mean Squared Errors (RMSE)				Sum of Squared Deviations		
	Full Equations	Constrained Equations	Difference	Percentage Difference	F Value of Difference in RSS (DF)	Critical F ($\alpha = .1$)	
PTO	0.7060	0.7043	0.0018	0.25%	-.085 (3,51)	5.15	
CSAT	0.0469	0.0479	-0.0010	-2.04%	0.714 (3, 51)	5.15	
WASG	0.1019	0.1010	0.0009	0.90%	-.223 (4, 50)	3.79	
PBIT/S	0.0089	0.0093	-0.0003	-3.54%	0.213 (14, 40)	1.89	

NB: Tests are reported for customary Granger causality models (i.e., column 2 of Table 6). Negative percentage differences indicate that full equations, which include hypothesized lagged independent variables, have lower RMSE and superior predictive ability.

TABLE 8**Second Qualitative Analysis of Comments Referring to Relations among DBSC Measures**

Relation codes	Company Managers					Distributors									Total
	J	K	L	M	N	A	B	C	D	E	F	G	H	I	
Cause & effect	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Finality	6	8	10	10	3	14	9	17	10	6	20	11	9	8	141
Logical	3	1	0	1	4	4	1	2	5	5	6	2	3	1	38
Total	9	9	10	11	7	18	10	19	15	11	26	13	12	9	179

Eighty-four of these comments referred to relations between specific pairs of measures.