The Effects of Total Quality Management on Corporate Performance: An Empirical Investigation

Author(s): George S. Easton and Sherry L. Jarrell

Source: *The Journal of Business*, Vol. 71, No. 2 (April 1998), pp. 253-307
Published by: The University of Chicago Press
Stable URL: https://www.jstor.org/stable/10.1086/209744
Accessed: 07-09-2018 17:51 UTC

George S. Easton
Emory University

Sherry L. Jarrell
Georgia State University

The Effects of Total Quality Management on Corporate Performance: An Empirical Investigation*

This article examines the impact of Total Quality Management (TQM) on the performance of 108 firms that began TQM implementation between 1981 and 1991. The impact of TQM is measured by comparing each firm's performance to a control benchmark designed to capture what the performance would have been without TQM. The findings indicate that performance, measured by both accounting variables and stock returns, is improved for the firms adopting TQM. The improvement is consistently stronger for firms with more advanced TQM systems. The possibility that downsizing could explain the improvement is also examined. The data do not support this hypothesis.

I. Introduction

The emergence of Total Quality Management (TQM) has been one of the most significant recent developments in U.S. management practice. The focus on the development of TQM systems in the U.S. appears to have begun around 1980 in response to global competition, primarily in U.S. manufacturing companies facing competition from Japan. By the mid- to late 1980s, the U.S. TQM movement had developed significant

* George S. Easton is associate professor of decision and information analysis, Goizueta Business School, Emory University, Atlanta, Ga. 30322. Sherry L. Jarrell is visiting assistant professor, Department of Finance, College of Business, Georgia State University, Atlanta, Ga. 30303. We would like to thank the companies that participated in the interviews. We would also like to thank the editor and two anonymous referees for comments that substantially improved the manuscript. In addition, we would like to thank the following for comments and suggestions: George Benson, George Benston, William Golomski, Mark Holder, Tom Hustad, Phillip Lederer, Albert Madansky, Harry Roberts, Vinod Singhal, and Marc Zenner. Finally, we would like to thank Melissa J. Smith and Karin Sparf Chill for excellent research assistance. This research was partially supported by National Science Foundation grants SBR-9523962 (Easton) and SBR-9523003 (Jarrell), by a grant from the conference ''Field Studies in Quality Management'' at the University of Rochester, Simon Graduate School of Business Administration, March 1993, and by summer research grants from the University of Chicago Graduate School of Business, and the Indiana University, School of Business.

(*Journal of Business*, 1998, vol. 71, no. 2)

0021-9398/98/7102-0004$02.50

momentum, in part because of the creation of the Malcolm Baldrige National Quality Award by Congress in 1987 and participation in the award by leading companies, such as AT & T, Motorola, Texas Instruments, Westinghouse, and Xerox.

Exactly what constitutes TQM is the subject of debate. In this article, we define TQM to be a management system that substantially addresses the criteria of the Malcolm Baldrige National Quality Award (National Institute of Standards and Technology [NIST] 1994). Although a complete definition of TQM is beyond the scope of this article, some of the key characteristics include[1]

- *Process focus.*—Process focus means an emphasis on the concept of process as a fundamental building block of the organization. This results in a widespread emphasis on process definition, process management, and process improvement.
- *Systematic improvement.*—Systematic improvement means a widespread systematic organizational focus on quality improvement, cycle-time reduction, and waste (cost) reduction and the adoption of a prevention-based orientation.
- *Companywide emphasis.*—The process concept and the emphasis on improvement are applied throughout the company, including to product development and business support processes.
- *Customer focus.*—Customer focus includes (1) emphasis on customer requirements and customer satisfaction to define product and service quality (''customer-defined quality''); (2) emphasis on customer service (lead-time reduction, on-time delivery, field support, technical support, etc.); (3) integration of customer information into the management and improvement systems—particularly into the new product development process and the production and service quality control and improvement processes; and (4) efforts to become integrated with customers as appropriate (often called ''partnering''), such as joint improvement teams, participation in the customer's new product development processes, or involving customers in the company's internal processes, such as planning, new product development, R&D, or technology forecasting.
- *Management-by-fact.*—Management-by-fact means an emphasis on deployment of systematic analysis and fact-based decision making driven by objective data and information. This includes a focus on deployment and tracking of metrics.
- *Employee involvement and development.*—Employee involvement in improvement (quality, cycle time, and waste), usually through teams, is widespread and there is a strong emphasis on employee

1. See Easton (1995) for a discussion of the characteristics of U.S. TQM.

development through training. This emphasis is generally associated with a tendency to drive decision making close to the actual processes and thus to a corresponding increase in employee empowerment.

- *Cross-functional management.*—There is explicit emphasis on cross-functional management that includes cross-functional improvement as well as cross-functional involvement in key processes such as new product development. Part of the cross-functional emphasis stems from the focus on processes (which typically cross multiple functions), although the emphasis is much stronger, recognizing cross-functional issues and involvement as requiring specific focus in order to achieve highly effective management systems.
- *Supplier performance and supplier relationships.*—Supplier management includes emphasis on supplier quality and service performance, supplier capabilities, supplier improvement, and supplier involvement and integration (supplier partnerships), such as joint quality improvement, and participation in new product development, technology development and planning, and even strategic planning.
- *Recognition of TQM as a critical competitive strategy.*—There is widespread recognition that implementation and aggressive refinement of the above management model is a critical competitive strategy and, thus, a primary concern of all levels of management, including senior management. The role of senior management in providing leadership for the development and deployment of quality management is a natural consequence of recognition of quality management as a critical competitive strategy.

There is considerable controversy concerning the effectiveness of TQM and research examining its impact is only beginning to emerge. Most of this research is based on cross-sectional surveys that examine the association between manager perceptions of the impact of TQM and model constructs based on questionnaire items that are intended to capture various aspects of the deployment of TQM. There is little empirical research that attempts to determine the impact of TQM on corporate performance by directly examining publicly available financial data. Of these studies, most do not focus directly on TQM but rather on related events, such as winning a quality award (e.g., Hendricks and Singhal 1996) or achieving International Organization for Standardization 9000 series (ISO 9000) registration (e.g., Anderson, Daly, and Johnson 1995). For a critical review of existing research on TQM that measures performance using publicly available financial data, see Easton and Jarrell (forthcoming).

This study examines the impact of TQM on financial performance for a sample of 108 firms. The study is based on a comprehensive

research methodology that combines (1) interview-based research to identify a sample of firms that have, in fact, made serious efforts to implement TQM systems in a majority of their business and (2) an empirical analysis of publicly available financial data using an improved benchmark and control methodology (Jarrell 1991) for isolating the impact of the adoption of TQM. We believe that the methodology developed in this study for examining the impact of a complex management phenomenon is an improvement over approaches generally taken in the literature and represents one of the contributions of this article.

II. Overview of the Research Methodology

The basic approach used in this study adapts the event study methodology, commonly used in empirical corporate finance, to examine the impact of TQM on firm financial performance. In this study, the ''event'' corresponds to the beginning of serious efforts to deploy a comprehensive TQM system. The impact of TQM is assessed by examining the unexpected changes in financial performance for a 5-year period following the beginning of deployment of the TQM system.

In most event studies in empirical finance, both the event and when it occurred can be unambiguously defined without much difficulty (e.g., the announcement of a merger). In this study, however, determining both whether and when an event has occurred is more difficult. First, whether or not a firm has seriously pursued TQM cannot be determined by relying on the firm's public pronouncements. Many firms claim to be implementing TQM when, in fact, they have made essentially no changes (other than in their public rhetoric). In other cases, TQM has been implemented in only a small fraction of their business. Second, firms seldom publicly announce the beginning of the deployment of their TQM systems. In fact, there is often no completely unambiguous start date. Rather, there is a period during which the firm's activities focus and efforts begin in earnest.

The lack of publicly available information about firms' implementations of TQM, the unreliability of their public statements, and the ambiguity of the start date of their TQM implementations are addressed in this study by interviews of a senior quality executive at each of the potential sample firms. Potential sample firms were first identified through public information sources as described in Section III. Interviews were then used to determine (1) whether a firm has, in fact, seriously pursued development of a TQM system; (2) the approximate extent of development and deployment of the firm's approaches; and (3) the approximate date that serious efforts began. The interviews were conducted by a former senior examiner for the Malcolm Baldrige Na-

tional Quality Award. The interview methodology is discussed in detail in Section IV.

The use of in-depth semistructured interviews to select the sample firms is an important difference between this study and typical studies based on questionnaires. The key reasons are that the interviews are interactive, flexible, and allow in-depth discussion and focused probing. This permits considerable verification of the information obtained and allows for clarification of terminology and adjustment for the specific knowledge and experience base of the interview subject. In addition, interviews conducted by an interviewer trained in evaluation against a TQM ''standard'' (the Baldrige Award criteria) allow external rather than respondent self-assessment of the company's TQM system against a well-developed operational definition.

In contrast, questionnaire-based approaches generally allow self-selection into the sample and rely on the managers' perceptions without critical evaluation. It is also very difficult in survey-based research to address the large variation in interpretation of terminology in different companies, and it is frequently unclear how respondents actually operationalize the questions. As a result, most questionnaire-based research is fairly superficial. These research issues are discussed further in Easton and Jarrell (forthcoming). Interview-based approaches, of course, also have disadvantages. These include dependence on the skill and knowledge of the interviewers, the difficulty of precise replication of the methodology, and the inability to examine the data collection instrument used.

In this study, interviews are also used to divide the sample firms into two groups based on the development of their TQM systems. The performance for these two groups is then compared. This provides an intrasample validation of the overall research method since if TQM positively affects performance, the more advanced firms should perform better than the less advanced firms.[2]

The event study approach is another important difference between this study and cross-sectional studies that examine association between performance and the reported use of various practices (e.g., employee participation). Such cross-sectional studies generally do not attempt to determine when the practices were initiated or to examine performance changes associated with actual implementation. The failure to focus specifically on performance changes associated with the actual changes in management practices greatly increases the possibility of confounding factors. Further, such studies provide weak evidence concerning

2. The validity of this comparison as an intrasample validation is discussed further in Section VI *G* 4. Because the decision to continue TQM implementation is endogenous, early financial success during the postevent period could influence the subsequent development of the firm's TQM system. The empirical analysis in Section VI *G* 4 indicates that this phenomenon does not drive the results of the comparison between the more advanced and less advanced firms.

causality, even when statistically significant associations are observed, because the direction of causality is often unclear. In many cases, it is at least as plausible that, because of the availability of additional resources, improved performance drives the more extensive use of the ''progressive'' practices typically examined in these studies as it is that the progressive practices caused the improved performance. While it is impossible to *prove* causality through observational studies (including this one), studies that focus as tightly as possible on the period of the management changes and that use a carefully developed control methodology clearly provide far more compelling evidence than those that do not.

The control methodology used to develop the performance measures is another critical research issue. To assess the impact of TQM, the company's actual performance would ideally be compared with what the performance would have been had the company not implemented TQM (i.e., a perfect ''clone'' but with no TQM). Since this is not possible, a benchmark performance measure must be constructed that, on average, captures what the performance would have been without TQM. In this study, performance is assessed using both accounting-based variables and daily stock returns over a 5-year period following the event. The performance measures are constructed somewhat differently for the accounting and stock return variables. For the accounting variables, the primary approach consists of two components. First, a firm's *unexpected performance* is measured by the difference between the firm's actual performance and an analyst's forecast made just prior to the event. Second, the event firm's unexpected performance is compared with the unexpected performance of a carefully matched control portfolio of three firms that do not appear to have implemented TQM. The control firms are matched to the event firm on the basis of industry, time period, analysts' projections of future performance, and, to the extent possible, on market size, debt-to-equity ratios, and a market risk factor. The impact on performance is then measured by the *excess unexpected performance*, the difference between the unexpected performance of the event firm and the unexpected performance of its control portfolio.

The use of analysts' forecasts in the accounting performance variables is important because such forecasts incorporate an expert's evaluation of the future impact of the firm's particular circumstances. It is these forecasts that allow the performance measure to adjust for firm-specific exogenous factors that are likely to affect future performance, including factors influencing the endogenous decision to implement TQM. The failure to control for such factors can introduce potentially significant bias into the results. Such factors may not be apparent in the firm's historical financial data (e.g., emerging foreign competition,

the expiration of a patent, developing labor issues, or pending regulatory or tax changes).[3]

The use of the control portfolios is also critical to correct for subsequent exogenous events during the postevent period (e.g., a recession). Since the control portfolio is also matched on the analyst's projection of future performance, the research design provides an additional control against systematic differences between the event and control firms in bias in the analysts' forecasts.[4]

The idea of assessing performance relative to a prediction of future performance (i.e., the unexpected performance) is a fundamental idea in financial theory. It is intrinsic in any analysis based on stock returns because stock prices are derived from the market's consensus forecast of expected future performance. Ideally, market consensus forecasts would have been used here. However, these are not directly available.[5] Analysts' forecasts, which represent an expert assessment, are used instead to proxy for market expectations.[6]

3. Adjusting for endogeneity of the decision to implement TQM (or any similar management decision) requires the prediction of what performance would have been had the same firms not implemented the changes. Thus, the performance measures used in this study must account for variables that are associated with the decision to implement TQM that would affect future performance even if TQM was not implemented. There are a variety of other approaches that might be taken other than the use of analysts' forecasts. For example, one might try to build econometric models that include exogenous variables thought to be associated with the choice to implement TQM. Actual performance of the event company could then be compared with the models' prediction. Forecasts based on such statistical models, however, have several disadvantages relative to analysts' forecasts (see n. 6 below). They are generally not developed on a firm-by-firm basis, do not incorporate information from sources other than the time series of accounting data, and are subject to errors caused by model building, outliers in the data, etc.

4. What is required is that the difference between the analysts' forecasts for the event firm and for the control portfolio is unbiased for the market consensus forecast of the expected difference. This means that the control portfolio methodology corrects for any systematic bias in the analysts' forecasts, provided such biases apply, on average, equally to the event and control firms (under the null hypothesis of no effect due to TQM).

5. The Institutional Brokers Estimate System market consensus forecasts are not used because they are limited to short-term earnings per share forecasts and are not time stamped in a way that allows reliable determination of when the forecasts were made relative to the event times. This study requires long-term forecasts and examines variables other than earnings per share.

6. In addition to the theoretical basis discussed here, there is also considerable empirical evidence that analysts' forecasts are effective proxies for market expectations (e.g., see Schipper 1991; Brown 1993). There is also evidence that analysts' forecasts are superior to time-series forecasts, at least for simple time-series models. Brown et al. (1987) find that the forecasting ability of Value Line analysts is superior to univariate time-series models. Brown and Rozeff (1978) compare the earnings predictions of Value Line analysts and forecasts in Standard and Poor's Earnings Forecaster to those from three different time-series models, including random walk and Box-Jenkins models, and find that Value Line analysts produce more small annual forecast errors and fewer large annual forecast errors. Finally, Brown et al. (1987), who examine Value Line forecasts, and Fried and Givoly (1982), who examine forecasts from Standard and Poor's Earnings Forecaster, find that 1-year-ahead analyst forecasts have a greater association with excess stock returns over the next year than do 1-year-ahead earnings forecasts made by time-series

In contrast to the methodology used here, traditional approaches in the empirical finance literature use preevent firm performance or postevent industry average performance as the control benchmark. Both of these approaches are unsatisfactory. The preevent performance benchmark fails to control for subsequent exogenous macroeconomic events. The postevent industry benchmark assumes that the firm, had it not adopted TQM, would have performed like the typical firm in the industry. This fails to address the endogeneity of the decision to implement TQM. Some more recent approaches are based on fitting structural models and comparing actual performance to the model's prediction (see, e.g., Healy, Palepu, and Ruback 1992). While superior to the preevent performance or postevent industry benchmark methods, such approaches generally assume that the structural equation is the same across event firms and is unaffected by subsequent exogenous events. They further assume that all of the factors likely to affect future performance, including those associated with the decision to implement TQM, are evident in the preevent financial performance data used to estimate the structural equation. Thus, they do not adequately control for bias due to endogeneity of the decision to implement TQM.[7]

The method developed in this article is used to assess performance of the TQM firms for appropriately scaled variables based on net income, operating income, and sales. Unfortunately, analysts' forecasts are not available for some other variables of interest (e.g., variables based on inventory levels or number of employees). For these variables, performance is measured by *excess actual performance*, the difference between the actual performances of the event firm and the control portfolio. While the evidence provided by these variables is much less compelling than when analysts' forecasts are available, we believe the results do contain some useful indications, particularly in the context of the overall analysis.

The impact of TQM is also evaluated using with-dividend continuously compounded daily stock returns. Because the stock price incorporates the market's forecast of a firm's future performance, it is not appropriate to use analysts' forecasts when examining stock returns. It is important, however, to control for the impact of postevent exogenous events. Thus, the performance measure for stock returns is based on the *excess actual returns*, the difference between the returns of the

models. These studies support the view that analyst expectations are a better proxy for market expectations than forecasts from time-series models. Section VI *G* 1 also provides empirical evidence validating the Value Line forecasts for the sample of event and control firms used in this study.

7. This remark also applies to the approach suggested by Barber and Lyon (1995). Their article examines several methods from event studies using accounting-based measures and concludes that test statistics are only well specified when sample firms are matched to control firms of similar preevent performance.

event firm and its control portfolio. For this measure to be valid, it is important that the event firms and control portfolios are well matched in terms of nondiversifiable risk. This is achieved as a result of the method of matching control firms discussed above, which includes consideration of expected future performance, market size, debt-to-equity ratios, and a market risk factor. Thus, the matched control portfolios control for both postevent exogenous events and nondiversifiable risk.

Despite the similarity in methodology, this study differs from typical empirical finance event studies in some important ways. First, this study does not focus on the effect of information events (''announcements'') on the capital market. While we examine stock returns, we use them for a different purpose—as a comparatively ''clean'' overall performance measure. Second, the event dates are not determined from public information but rather from private information obtained through interviews. Third, the phenomenon of interest (the deployment of TQM) does not occur at a discrete point in time like a typical announcement but rather occurs over a period of at least several years. Thus, we do not expect stock price reactions around event time 0. Instead, we examine a 5-year period following the beginning of the implementation of TQM. While some of the benefits of TQM, such as certain types of cost reductions, can be obtained relatively quickly, many others, such as improvements in new product development or increased market share because of increased customer satisfaction, require at least several years to become evident in the firm's accounting data. Many benefits of TQM may even occur after the 5-year postevent period that we examine. Further, during the period that the firms in this sample began implementing TQM, the capital markets had little basis for assessing TQM's impact; the evidence is only now beginning to emerge. Thus, it is not unreasonable to expect that the impact on stock return performance will occur throughout the 5-year postevent period as the results (positive and negative) of TQM implementation accrue and become evident in the firms' accounting data.

In interpreting this study, it is important to understand that we are attempting to examine transient performance effects due to introduction of a new management ''technology.'' In a theoretical setting where managers always instantaneously select the optimal strategy for maximizing firm performance on the basis of the available information set, a firm's decision to implement TQM or not would always be deterministically driven as the optimal response to exogenous variables. After controlling for all of these exogenous variables, there would be no observable effect due to TQM. This theoretical argument is not unique to TQM; it also applies to other management decisions, including restructuring decisions (e.g., mergers), which are frequently the focus of similar research examining their performance impact.

These assumptions, of course, are unrealistically strong. Managers

do not always make optimal decisions and certainly do not always do so instantaneously. What managers seek are strategies for moving their companies toward a dynamically changing optimum. Potential strategies include implementation of TQM (among many others—including restructuring). It is of interest to examine whether or not such strategies generate value for the companies that implement them. The performance impact can be examined only because of deviations from the theoretical setting described above. These deviations should be transient as competitive pressure drives the economic system toward optimality.

III. Data Sources and Sample Design

Candidate event firms were initially identified through publicly available information sources. The search was intended to be comprehensive, but not exhaustive. The primary sources were the database of online full-text annual reports (the ARS database) from Nexus/Lexus (since 1987), the Businesswire full-text database of press releases (since 1986), Standard and Poor's Corporate Register of Directors and Executives (1993), and the list of Baldrige Award site-visited companies in the report by the U.S. General Accounting Office (1991) on quality practices.

The study was conducted in two phases. The pilot phase was based on annual report searches for the key words ''total quality management,'' ''just-in-time,'' ''JIT,'' ''Baldrige,'' ''Deming,'' ''Juran,'' and ''Crosby.'' These searches identified 274 firms. Relevant excerpts were then reviewed to select only the firms whose annual reports clearly indicated implementation of at least one specific quality management approach (e.g., statistical process control [SPC], just-in-time [JIT], quality training, improvement teams, etc.)

Review of the annual report searches resulted in a list of 78 firms. These firms were contacted to set up an interview with a senior manager familiar with the development of the firm's quality management systems. Of the 78 firms, 59 were interviewed. In 11 cases, firms were not interviewed because it became clear in trying to set up the interview that the firm was not actually implementing TQM. In the remaining eight cases, the request was refused. Of the 59 firms interviewed, 15 were eliminated because the interviews did not indicate serious efforts to implement TQM in a majority of their business. An additional five firms were eliminated because the required performance data were not available. The remaining 39 firms formed the pilot sample. Interviews for the pilot sample were conducted between January and March 1993.

In the second phase of the study, additional candidate firms were sought from a variety of sources. First, an additional 54 firms were selected after a second review of the original 274 firms identified

through the annual report searches. Second, new firms were sought through additional searches and sources. The Businesswire database was searched for references to quality awards. Searches were also made for quality-related executive titles. The annual report database was searched for ''quality'' within five words of ''vice president'' or ''director,'' and the Businesswire database was searched for ''total'' or ''continuous'' within three words and ''focus'' and ''satisfaction'' within five words of ''vice president'' or ''director.'' These searches identified 89 firms. The 1993 Standard and Poor's Corporate Register of Directors and Executives was searched for ''quality'' within five words of ''vice president'' or ''director,'' which identified 71 firms. Finally, lists of site-visited firms from the GAO study and lists of the institutional affiliations of Baldrige Award examiners for the years 1989–93 (available from the Baldrige Award office) were reviewed, which identified 67 additional firms. Thus, 281 new candidate firms were identified in the second phase of the study.

As in the pilot phase of the study, the information on these firms was reviewed for evidence of specific quality management approaches. This resulted in a list of 129 firms that were then contacted for interviews. Of these, 117 agreed to be interviewed. Of the 12 firms that were not interviewed, six declined to participate, and six obviously did not have TQM programs. Of the firms that were interviewed, 38 were eliminated because the interviews did not confirm serious efforts to implement TQM in a majority of the business, and 10 were eliminated because the required performance data were not available. This resulted in 69 additional event firms. These interviews were conducted between August 1993 and January 1994.

In summary, information on over 500 firms was reviewed to identify potential sample firms. Of these, a total of 207 were approached for interviews. Fourteen firms declined to participate, which gave an overall response rate of 93%. In trying to set up the interviews, 17 firms were determined not to have a TQM system. A total of 176 firms were actually interviewed, and 53 firms (30% of those interviewed) were eliminated because their efforts to implement TQM did not appear to be adequate. An additional 15 firms were eliminated because the required performance data were unavailable. This process resulted in 108 event firms in the final sample (see app. A, table A1).

The Value Line Investment Survey was used as the source of analysts' forecasts, the primary source of the accounting data, and to select the control firms (see Sect. V). For the measures based on the Value Line analysts' forecasts, performance is examined for year 1, year 2, and the average of years 3–5 following the event. Long-term data were available for 100 of the 108 events. The Compustat database compiled by Standard and Poor's was used for data that Value Line does not report (inventory levels and the number of employees). Daily

stock returns were obtained from the database compiled by the Center for Research in Security Prices (CRSP) at the University of Chicago.

IV. Interview Methods

Each candidate sample firm was contacted, first by a letter describing the project and then by telephone, to set up an interview with a senior manager familiar with the development of the company's quality management systems (generally a vice president or director of quality). The interviews generally lasted about 45 minutes and were conducted by George S. Easton, a former senior examiner for the Malcolm Baldrige National Quality Award. The objective of the interview process was to develop a time line of the development of the company's TQM systems, to determine what key approaches were used, and to assess the actual extent of deployment through in-depth probing in a few areas. The interviews were semistructured and allowed flexibility in the topics discussed. The managers were promised complete confidentiality concerning the interview content.

The interview process occurred in two phases. The objective of the first phase was to elicit from the manager, with minimal prompting, the major milestones in the development of the company's TQM approaches. Questions were asked as necessary to establish the level of detail desired and to determine as specifically as possible the dates of the events surrounding the beginning of the TQM approaches. Questions about approaches or methods not mentioned by the manager were avoided in order not to influence the manager's description. These impromptu descriptions are very revealing about the aspects of the development of the TQM system that the manager believes are important and what the key drivers of the company's system actually are; that is, how the company ''thinks'' about its TQM systems.

The second phase of the interview process was intended to fill in any important gaps and to probe some key areas in order to assess actual levels of deployment. The list of interview topics given in appendix B was used to prompt the interviewer. The objective was not to discuss every topic but rather to discuss in detail a few areas as appropriate for the company's approaches and the expertise and experience of the manager being interviewed. If not adequately addressed by the initial description of the time line, four areas were always covered: production, customer satisfaction measurement, supplier management, and new product development and design. In general, the extent of deployment of approaches mentioned was assessed by asking specific questions concerning the number of employees involved, the training they have received, and the dates of the various events mentioned. Other questions used to determine the actual extent of deployment in-

cluded asking what were the most important barriers to implementation, what should have been done differently if this approach was implemented again, what lessons were learned, and what changes or improvements have been made since the initial approach. When the approaches described have actually been deployed, there is generally a rich ''story'' surrounding them, and it is fairly easy for a knowledgeable interviewer to determine if significant deployment has actually occurred.

Companies were included in the sample if, on the basis of the interviews, they appeared to have made serious efforts to implement TQM approaches in the majority of their business. Deployment must have been in a majority of the company (as measured by sales) in order for there to be any reasonable expectation that the results could be observable in the company's overall financial data. The standard of ''serious efforts'' for inclusion in the sample is quite low; it is not a requirement that the company's efforts resulted in a comprehensive and well-integrated approach.

Companies were eliminated from the sample for a variety of reasons. In many cases, the reason was that the TQM efforts were deployed in only a small fraction of the company. Other reasons ranged from a lack of evidence of any significant deployment efforts to confusion of TQM with other approaches (such as quality improvements solely due to automation).

The start dates for the sample companies were chosen, on the basis of the time line developed from the interview notes, to be about 6 months after the beginning of the first major initiative. This initiative was usually the deployment of widespread quality training. In some companies, however, other initiatives marked the beginning of their TQM systems, such as major changes in customer satisfaction measurement or new product development, widespread deployment of SPC, or deployment of a quality management systems assessment process (e.g., Baldrige-based assessment). The start date was chosen to be 6 months after the beginning of the first major initiative because most such initiatives take substantial time to roll out. For example, it is not uncommon for widespread training initiatives to take over a year to complete in a large company, and there is usually an additional lag before substantive organizational or operational changes occur.

The start date determines when the analysts' forecasts were made that are used in the performance measures. In order for the difference between the actual postevent performance and the analyst's forecast to capture the unexpected performance due to TQM, the forecast should be made prior to the analyst's incorporating knowledge about the firm's TQM initiatives. This suggests selecting an early event time 0 to ensure that the analyst is not aware of the TQM initiative. However, too early an event time 0 truncates the postevent period, which is limited to

5 years, and may result in a failure to capture the main performance improvement due to TQM. The selection of event time 0 to be 6 months after the initial deployment of the first major initiative compromises between these conflicting objectives.

Despite the fact that event time 0 was selected to be 6 months after the beginning of the first major initiative, there are several reasons that it is unlikely that the analysts' forecasts are affected. They include the following: (1) there is almost never any public information available about the initiatives until later than this period; (2) any claims made by management about their intentions contain little substantive information about whether or not serious efforts to implement TQM will actually be made; (3) during the period we are studying, TQM was new and thus there would be little or no basis for updating the forecasts; and (4) the texts accompanying the Value Line analysts' forecasts were also reviewed, and in no case was there any mention of the firm beginning a quality-related initiative. More important, however, any such leakage into the analysts' forecasts biases against finding an effect due to TQM and thus makes the results of this study conservative.

The companies that were retained for the sample were divided into a group of 44 firms with more advanced TQM systems and a group of 64 firms with less advanced TQM systems by making a rough estimate, based solely on the interviews, of what the firm's score would be in terms of the approach and deployment (not results) areas of the Baldrige criteria. It should be noted that the interviews focused only on the approaches taken and the extent of their deployment and not on operational or financial results. The firms selected as more advanced had estimated scores above 450 out of 1,000 possible points. This represents considerable success developing and deploying a TQM system. The median score of companies that apply for the Baldrige Award is generally below 500.

The key differences between the more advanced and less advanced firms were in the scope of the issues addressed by their TQM systems and the extent of deployment of their approaches. Some of the firms in the less advanced group had successfully deployed basic approaches, such as quality training and improvement teams, but had then not further developed their quality management systems. Others had developed approaches which address a broader scope of the Baldrige criteria but had only limited deployment. In contrast, firms in the more advanced group had better deployment of the basic approaches together with the deployment of a broader scope of systems. These companies typically have had multiple major phases in the development of their TQM systems whereas less advanced firms typically have only completed one major phase. For example, a typical advanced firm might start with an initial phase focused internally on SPC and quality im-

provement teams, followed by a second phase that might focus on design quality, internal self-assessment, or customer satisfaction measurement (and feedback of such information into the company's internal processes). Subsequent phases would then focus on approaches and issues not already addressed.

V. Analysis Methods

This study examines statistical evidence against the null hypothesis that implementation of TQM does not improve corporate performance. Performance is measured by accounting variables, primarily focusing on net income, operating income, sales, and inventory, and by continuously compounded with-dividend stock returns. As discussed in Section II, in order to be convincing, the performance measures must (1) take firm-specific factors into account, including those associated with the (endogenous) decision to implement TQM and (2) compensate for postevent macroeconomic or industry-specific developments that are likely to affect firm performance. The approach used here was developed by Jarrell (1991) to address these issues.

A. Control Portfolio Selection

All of the performance measures examined rely on matched control portfolios. For each event firm, a control portfolio of three firms that do not appear to have implemented TQM is formed by matching them to the TQM firm on the basis of industry, calendar time, projected performance, and, to the extent possible, market size, debt-to-equity ratios, and a market risk factor (the Value Line ''safety'' ranking).[8] Matching on industry and calendar time is designed to control for various economic and regulatory influences. The industry classifications are defined by the Value Line Investment Survey and verified with the standard industrial classification (SIC) code. The matching included a detailed review of product lines as described by Value Line, so the matching realized is substantially better than that provided by using the Value Line industry classifications or SIC codes alone.

The ''projected performance'' matching of the control firms to the event firms is based on the Value Line ''timeliness'' rank. The timeliness rank summarizes the analyst's assessment of the firm's expected

8. Eighteen control firms were interviewed to provide some verification that the control firms have not made significant efforts to implement TQM. All were determined to be appropriate controls. It would be impractical to interview the entire control sample. In addition, especially for larger firms, failure to detect TQM-related efforts through the searches performed provides considerable evidence that these firms have not made significant TQM efforts. More important, however, contamination of the control portfolio by firms that have implemented TQM should bias the results against finding an effect associated with TQM.

stock price performance over the next 12 months relative to the other firms covered by Value Line.[9] Whenever possible, firms were selected whose timeliness rank at the time of the event differed by no more than one from the event firm rank. These firms were then narrowed to three control firms per event, first by choosing those closest in size to the TQM firm and then (if more than three remain) those whose debt-to-equity ratio and Value Line safety rank are closest to the TQM firm. Size is measured by total debt plus the market value of equity and preferred stock as reported by Value Line during event year 0.

Matching on the basis of the timeliness rank incorporates into the control portfolios as much information as possible about the expected performance of the TQM firms. Because the control firms are selected to have an outlook similar to that of the event firm, such matching has the potential to control for effects such as systematic differences in forecast accuracy between firms that are forecast to perform very well and those that are forecast to perform poorly. The existence of this type of bias is plausible. For example, because of the phenomenon of regression toward the mean, analysts' forecasts may be systematically too high for firms that are expected to perform very well and systematically too low for firms that are expected to perform poorly. In such a case, failure to control for projected performance could introduce bias into the results, especially if the sample has a high concentration of firms that are expected to either perform very well or very poorly.[10] In a similar manner, matching that considers the timeliness rank also minimizes the effects of any systematic differences in responses to subsequent economic events between firms with very positive and very negative outlooks.[11]

9. The timeliness rank is scaled from 1–5 (1 corresponds to the highest projected performance) and is updated approximately every quarter. Value Line indicates that the rank is based on three criteria: (1) the firm's industry-adjusted price-to-earnings ratio from the previous 12 months relative to the last 10 years, (2) the year-to-year change in the quarterly earnings of the stock compared with that of all Value Line stocks, and (3) an earnings ''surprise'' factor.

10. Analysis of the forecasts for the control firms (which, with respect to this issue, are not contaminated by the effects of implementing TQM) does show a very slight but not statistically significant pattern of overestimation for firms with low timeliness ranks (strong expected performance) and underestimation of performance for firms with high timeliness ranks (poor expected performance). This, however, should not be intepreted as a justification for not using the analysts' forecasts since the bias associated with the high and low timeliness ranks is very small and is not statistically significant. Further, it is much smaller than would occur when matching to firms with very different timeliness ranks and is corrected by the matching strategy used here. Finally, the average timeliness rank of the event firms is 3.08, so there is no concentration of firms in the sample with very high or very low timeliness ranks.

11. The following specific example further illustrates these issues. The event firm United Technologies (UT), a diversified company, has a TQM starting year of 1984. At that time, UT's timeliness rank was 1, the highest rank. According to the analyst, ''The key here is technological integration. Unlike the typical conglomerate, UT isn't simply equal to the

B. Accounting Variables: The Primary Performance Benchmark Method

The primary performance benchmark compares each event firm's and corresponding control firm's performance with Value Line analysts' forecasts made prior to the event. Specifically, for event firm *i* for postevent year *t*, the unexpected performance $U_i^E(t)$ is

$$U_i^E(t) = P_i^E(t) - F_i^E(t),$$

where $P_i^E(t)$ is the actual performance of TQM firm *i* for postevent year *t* and $F_i^E(t)$ is the Value Line analyst's forecast of that performance made prior to the event. The unexpected performance $U_i^{Cj}(t)$ for the firms in the control portfolio is similarly defined:

$$U_i^{Cj}(t) = P_i^{Cj}(t) - F_i^{Cj}(t),$$

where $P_i^{Cj}(t)$ is the actual performance and $F_i^{Cj}(t)$ is the forecast performance for period *t* for control firm *j* corresponding to event *i*. The unexpected performance $\overline{U}_i^C(t)$ for the control portfolio is the average of the unexpected performance for the three control firms: $\overline{U}_i^C(t) = {}^1/_3 \sum_{j=1}^{3} U_i^{Cj}(t)$.

For the accounting variables, the primary measure examined for evidence of the impact of TQM on firm performance is the excess unexpected performance $XU_i(t)$, the difference between the unexpected performance of the event firm and the unexpected performance of the corresponding control portfolio. Thus,

$$XU_i(t) = U_i^E(t) - \overline{U}_i^C(t).$$

---

sum of various unrelated parts.'' The analyst goes on to say that ''whatever it may have once been, UT is now a large and diverse company with a clear business strategy.'' If the timeliness rank is ignored in matching the controls, ITT (timeliness rank 4) would replace Kaman (timeliness rank 2). While ITT is a better match than Kaman in terms of size, it is far worse in terms of the analyst's assessment of future prospects. Specifically, ITT's ''top priority is to become a major force in telecommunications, through its new digital telephone switch, the System 12.'' While the ''System 12 is succeeding overseas,'' it is ''incompatible with the current generation of equipment used here.'' The analyst concludes that ''success in this country will not come overnight,'' that ''the telecommunications thrust is expensive,'' and that ITT ''will probably step up its divestiture activity.'' In contrast, for Kaman ''1984 is shaping up as a record year'' as a result of ''record gains from both the diversified and industrial distribution divisions.'' For example, examining the operating margin, UT and all three of the original control firms (including Kaman) ended up performing worse than their forecasts for years 3–5. United Technologies' performance was about 2% worse than the average of the control portfolio. International Telephone and Telegraph's performance, however, was substantially better than forecast (unlike any of the three original control firms), so the substitution of ITT for Kaman would make the performance of UT appear worse (2.5% worse than the control portfolio instead of 2% worse). While this result is anecdotal and thus may be due merely to random variation, it is also possible that the conditions that resulted in ITT's poor timeliness rank make it a poor control for the forecast error of a firm expected to perform very well.

Value Line analysts' forecasts are given for 1 year ahead, 2 years ahead, and the average of years 3–5 ahead. Thus, the excess unexpected performance cannot be calculated separately for postevent years 3, 4, and 5. Instead, following the analysts' forecasts, the long-term performance measure is based on unexpected average performance for years 3–5. Specifically, for event firm i, the unexpected average annual performance $\overline{U}_i^{\mathrm{E}}(3{:}5)$ for postevent years 3–5 is

$$\overline{U}_i^{\mathrm{E}}(3{:}5) = \overline{P}_i^{\mathrm{E}}(3{:}5) - \overline{F}_i^{\mathrm{E}}(3{:}5),$$

where $\overline{P}_i^{\mathrm{E}}(3{:}5) = [P_i^{\mathrm{E}}(3) + P_i^{\mathrm{E}}(4) + P_i^{\mathrm{E}}(5)]/3$ and $\overline{F}_i^{\mathrm{E}}(3{:}5)$ is the analyst's forecast of the average annual performance over years 3–5. Unexpected average performance for years 3–5 is calculated for each control firm in the same manner. Paralleling the development above, the excess unexpected average performance for years 3–5 is then calculated.

C. Accounting Variables: When Forecasts Are Unavailable

Value Line analysts' forecasts are not available for variables based on inventory levels and the number of employees. For these variables, excess actual performance is examined.[12] The excess actual performance $X_i(t)$ for event i for postevent year t is

$$X_i(t) = P_i^{\mathrm{E}}(t) - \overline{P}_i^{\mathrm{C}}(t),$$

where $\overline{P}_i^{\mathrm{C}}(t) = \frac{1}{3}\sum_{j=1}^{3} P_i^{\mathrm{C}j}(t)$. Note that $P_i^{\mathrm{E}}(t)$ and $P_i^{\mathrm{C}j}(t)$ are the actual performances for the event and control firms as defined in Section V*B*. In addition, excess average performance for both a 5-year preevent and a 5-year postevent period is examined; that is,

$$\overline{X}_i^{\mathrm{pre}} = \frac{1}{5}\sum_{t=-5}^{-1} P_i^{\mathrm{E}}(t) - \frac{1}{5}\sum_{t=-5}^{-1} \overline{P}_i^{\mathrm{C}}(t),$$

and

$$\overline{X}_i^{\mathrm{post}} = \frac{1}{5}\sum_{t=1}^{5} P_i^{\mathrm{E}}(t) - \frac{1}{5}\sum_{t=1}^{5} \overline{P}_i^{\mathrm{C}}(t).$$

Finally, the difference between the postevent and preevent excess average performance is also examined:

$$D_i = \overline{X}_i^{\mathrm{post}} - \overline{X}_i^{\mathrm{pre}}.$$

12. As discussed in Sec. II above, we believe the validity of analysis based on excess actual performance is considerably weaker than that based on excess unexpected performance. We nevertheless examine these variables because of their close link to the methods of TQM. In the context of the other analysis presented, we believe these variables do provide some useful indications.

D. Stock Returns

The impact of TQM is also examined using continuously compounded with-dividend daily stock returns. As discussed in Section II, it is not appropriate to use analysts' forecasts in conjunction with stock returns. Thus, the stock return performance measure is the *excess cumulative daily return.* The excess cumulative daily return $XCR_i(t)$ for event i at postevent day t is

$$XCR_i(t) = CR_i^{E}(t) - \overline{CR}_i^{C}(t),$$

where $CR_i^{E}(t)$ is the postevent cumulative daily return at day t for event firm i and $\overline{CR}_i^{C}(t)$ is the average of the cumulative returns at day t for firms in the corresponding control portfolio. Thus, $\overline{CR}_i^{C}(t) = \frac{1}{3}\sum_{j=1}^{3} CR_i^{Cj}(t)$, where $CR_i^{Cj}(t)$ is the postevent cumulative return at day t for firm j of the control portfolio C for event i. The cumulative returns $CR(t)$ are defined similarly for the event and control firms: $CR(t) = \sum_{t'=1}^{t} r(t')$, where $r(t')$ is the with-dividend continuously compounded daily stock return for day t' following event time 0. The excess average monthly stock returns for the preevent and postevent periods and the difference of the differences are also examined, where the monthly returns are calculated by cumulating the continuously compounded with-dividend daily returns for the month.

As described above, the control portfolio methodology matches control and event firms as closely as possible, with the result that the event and control firms are closely matched on systematic risk. However, the stock returns analysis was also repeated using beta excess returns. The results, which are not presented here, are similar, indicating that differences in systematic risk between the events and controls are not driving the stock return results.

VI. Results

This section describes the results for both the accounting measures and stock returns. All results are for either excess unexpected or excess actual performance (depending on the availability of analysts' forecasts) of the TQM firm in comparison with the matched non-TQM control portfolio. The analysis of the accounting variables focuses on net income, operating income, and inventory scaled by measures of firm size based on sales, assets, or number of employees. Results for sales to assets are also presented. Results are given for the full sample of 108 TQM events and for the subsamples of event firms with more advanced and less advanced TQM systems. The analysis is repeated using only the 93 manufacturing firms. We examine manufacturing firms separately because the early development of U.S. TQM was pri-

TABLE 1 Distribution of the Year of Implementation of TQM

Year of Implementation	No. of TQM Firms
1981	1
1982	3
1983	9
1984	11
1985	7
1986	15
1987	17
1988	10
1989	17
1990	15
1991	3
Total	108

marily in manufacturing companies, and as a result, the methods of TQM are better developed in this context.

Summary statistics for the full sample of 108 TQM event firms are given in tables 1–3. Event year 0, the beginning of the firm's TQM implementation, spans the years 1981–91 (table 1). The sample firms represent 32 different industries (table 2) and range in market size from $76 million to $73 billion, with a mean of $5.4 billion (table 3).

TABLE 2 Distribution of the TQM Firms by Industry

Industry	No. of Firms	Industry	No. of Firms
Aerospace	4	Machine tools	1
Air transport	1	Machinery	5
Auto and truck	3	Medical supplies	2
Auto parts	3	Metals and mining	1
Banking	2	Office equipment and supplies	4
Building materials	1	Oil field services	1
Chemicals	12	Packaging and containers	2
Computers and peripherals	10	Paper and forest products	5
Diversified	3	Petroleum	1
Electric utilities	2	Precision instruments	3
Electronics/electrical equipment	12	Publishing	1
Financial services	1	Semiconductors	11
Food processing	1	Steel	1
Furniture and home furnishings	1	Telecommunications	4
Home appliances	2	Tire and rubber	2
Household products	1	Trucking and transport leasing	5
		Total	108

TABLE 3 TQM Firm and Control Firm Market Size (in Billions $)

Market Size	TQM Firms	Non-TQM Control Firms
Mean size	5.4	2.4
Median size	1.5	.9
Minimum	.0759	.0476
Maximum	73.3	35.0

SOURCE.—Both the debt and equity variables are taken from the Value Line Investment Survey published during event year 0.

NOTE.—Market size is the market value of equity (including preferred stock) plus total debt.

A. Statistical Analysis Methods

All tables, except for tables 10 and 19, report the medians of the performance measures. The medians are used because, especially for the accounting variables, the data are not normally distributed. Deviations from the normal distribution include the presence of outliers, wide tails, and, for some variables, skewness. Medians are extremely robust to these types of problems. The tables for the accounting variables also present sign test *p*-values, which test the one-sided null hypothesis that TQM does not improve performance against the alternative that performance is improved. Sign tests were used because they are nonparametric and thus robust to the kinds of deviations from the normal distribution described above. Wilcoxon signed-rank tests were not used because this test assumes that the distribution of the data is symmetric, an assumption violated by the accounting data. When there is skewness, the test is not valid and can be inconsistent with the actual medians (e.g., the sample median can be negative, while the Wilcoxon test indicates that the center of symmetry of the distribution is positive).

The tables also compare the results for the less advanced and more advanced TQM firms. Wilcoxon rank-sum tests were used to test the null hypothesis that the distribution of the performance measure for the more advanced firms is not stochastically larger than that of the less advanced firms (against the alternative that it is stochastically larger). A distribution $F(x)$ is stochastically larger than a distribution $G(x)$ if $F(x) \geq G(x)$ for all x, but $F(x) \neq G(x)$. Unlike the Wilcoxon signed-rank test, the Wilcoxon rank-sum test does not assume that the distributions are symmetric or that, under the null hypothesis, the two distributions are identical. Wilcoxon rank-sum tests are also used for several other comparisons between subsamples. The specific hypotheses are described in the table legends.

For the cumulative daily stock return data in tables 9, 12, 18, and 21, Wilcoxon signed-rank tests are used rather than sign tests. The rea-

sons are that (1) normal probability plots do not indicate departure from symmetry; (2) there is much empirical evidence indicating that daily stock returns are reasonably close to normally distributed; and (3) Wilcoxon signed-rank tests are more powerful than sign tests when the underlying distribution is symmetric.

Tables 10 and 19 present analysis of the cumulative daily stock return data based on the assumption that the stock returns are multivariate normal. Thus, sample means are reported. The *p*-values in these tables use test statistics that are based on estimates of the variances and covariances between firms whose event year 0 is the same calendar year. Thus, comparison of these tables with tables 9 and 18 allows an assessment of the impact of any correlation due to industry and event-year clustering.

B. Accounting Variables: Excess Unexpected Performance

Table 4 shows the results for excess unexpected performance for net income to sales (NI/S), net income to assets (NI/A), operating income to sales (OI/S), operating income to assets (OI/A), and sales to assets (S/A). For the full sample of firms, the table shows that the median excess performance is positive for all of the variables for the average of years 3–5 indicating that, for all of the variables examined, more than half of the event firms performed better in comparison with the analysts' forecasts than did the matched control portfolios. This improvement for years 3–5 is significant at the 1% level for OI/A, at the 5% level for NI/S and NI/A, and at the 10% level for S/A. Performance is also improved for all the variables in years 1 and 2 except for OI/S and OI/A, where there is a decline in performance in year 2. While the results for years 1 and 2 are generally not statistically significant, this provides some evidence against the idea that implementing TQM hurts short-term performance. Note that the improvement is much larger for years 3–5 than for years 1 and 2 for all of the variables except OI/S, which is consistent with the hypothesis that the most important impact of implementing TQM is on longer-term performance.

For the firms with more advanced TQM systems, except for OI/S, the results for years 3–5 are uniformly better than for the firms with less well developed systems. For the more advanced firms, in spite of the much smaller sample size ($N = 44$), the years 3–5 performance improvement is significant at the 1% level for OI/A and S/A, at the 5% level for NI/A, and at the 10% level for NI/S. The results are also better for the years 3–5 performance than for the year 1 and year 2 performance for all variables except OI/S. For the more advanced firms, the improvement in year 1 is significant for all of the variables and significant in year 2 at the 10% level for NI/S. For the less advanced firms, while all of the medians for years 3–5 are positive, none of the variables are significant. There is no indication that short-term perfor-

TABLE 4 Excess Unexpected Performance for the Accounting Variables

	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
Variable and Event Year	Median (%)	*p*-sgn	*N*	Median (%)	*p*-sgn	*N*	Median (%)	*p*-sgn	*N*	*p*-wrs
NI/S:										
1	.26	.11	108	−.03	.55	64	.54	.03	44	.04
2	.25	.19	108	−.03	.55	64	.54	.09	44	.09
3–5	.60	.03	100	.47	.11	56	1.12	.09	44	.12
NI/A:										
1	.39	.11	108	−.04	.65	64	.83	.01	44	.02
2	.49	.07	108	.40	.13	64	.52	.23	44	.22
3–5	.91	.03	100	.37	.17	56	1.86	.05	44	.06
OI/S:										
1	.03	.50	108	−.43	.92	64	.58	.06	44	.11
2	−.12	.78	108	−.47	.92	64	.10	.38	44	.10
3–5	.04	.46	100	.16	.45	56	−.01	.56	44	.30
OI/A:										
1	.46	.11	108	−.17	.73	64	1.27	.01	44	.04
2	−.12	.61	108	.02	.55	64	−.12	.67	44	.54
3–5	1.52	.01	100	.37	.17	56	2.98	.01	44	.04
S/A:										
1	1.72	.05	108	.46	.35	64	3.86	.03	44	.10
2	1.17	.25	108	.05	.55	64	2.82	.15	44	.36
3–5	4.89	.07	100	.03	.55	56	8.40	.01	44	.01

SOURCE.—The data and forecasts used to construct the performance measures were obtained from the Value Line Investment Survey.

NOTE.—NI/S = net income to sales, NI/A = net income to assets, OI/S = operating income to sales, OI/A = operation income to assets, and S/A = sales to assets. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p*-sgn contain *p*-values for the one-sided sign test of the null hypothesis H_0: true median ≤ 0 against the alternative H_A: true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less advanced firms.

mance is improved for the less advanced firms, and in fact, there may be some evidence that it declines for OI/S.

The *p*-values for the Wilcoxon rank-sum test indicate that the improvement for the years 3–5 performance for the more advanced firms in comparison with the less advanced firms is significant at the 1% level for S/A, the 5% level for OI/A, and the 10% level for NI/A. The difference in year 1 performance is significant at the 5% level for NI/S, NI/A, and OI/A and at the 10% level for S/A. The difference in year 2 performance is significant at the 10% level for NI/S and OI/S.

In summary, table 4 provides strong evidence of overall improvement in longer-term performance for these accounting variables for the full sample of TQM events. This improvement is stronger for the more advanced firms and weaker for the less advanced firms, with performance for the less advanced firms, for the most part, not statistically

different from the controls. The longer-term performance is stronger than the short-term performance. For the full sample, there is no evidence that short-term performance is hurt by the implementation of TQM. There is, in fact, evidence that even the short-term performance for the more advanced TQM firms is improved.

C. Accounting Variables: Excess Actual Performance

Table 5 shows excess actual net income per employee (NI/E) and operating income per employee (OI/E) for a period from 5 years before the event to 5 years after the event. For the full sample, there appears to be an overall declining trend in NI/E preceding the event and a generally improving trend following the event, resulting in positive (but not significant) median performance for year 5. At the bottom of each panel, the table also shows the excess performance for the average of years −1 to −5, the excess performance for the average of years 1–5, and the difference between the preevent and postevent period excess average performance. The median average performance is negative for both the preevent and postevent periods. The median difference between the preevent and postevent period averages is positive but not statistically significant.

For OI/E, the performance varies around zero through year 1 following the event. For years 2–5, there is an improving trend with significant positive performance at the 5% level in years 3 and 5. Note that, because of the availability of the data, the sample size decreases over years 2–5, making it more difficult to obtain statistical significance for the longer-term data. For year 5, the median improvement in operating income for the event firms in comparison with the control firms is approximately $3,000 per employee. The median postevent period average is positive and statistically significant at the 10% level. The median difference in the postevent and preevent averages is positive, but not significant.

For the more advanced TQM firms, a similar but stronger pattern emerges. The median excess NI/E turns positive in year 2 and, with a *p*-value of .11, is almost significant at the 10% level in year 5. The median excess NI/E in year 5 is approximately $840. The median postevent period average is positive but not significant. The median difference in the preevent and postevent averages, however, is positive and significant at the 10% level, providing some evidence of improvement between the preevent and postevent periods. For OI/E, the performance is significantly positive at the 5% or 10% level for years 2–5 following the event with a median excess OI/E of $4,830 in year 5. The median postevent period average is about $3,680 and is significant at the 1% level. The median difference in the postevent and preevent averages is about $2,630 and is significant at the 10% level. In contrast, for the less advanced firms, excess NI/E is negative, although improving,

throughout the postevent period. The excess OI/E fluctuates around zero and is not statistically significant. For both of these variables, the median preevent and postevent averages and their median differences are not significant. The postevent average excess performance and the differences in the preevent and postevent averages are greater for the more advanced than for the less advanced firms with significance at the 3% level or better for both variables.

Table 6 shows the excess actual sales per employee (S/E). For the full sample, the median excess S/E during years 3–5 is negative ranging from −$1,890 to −$5,900. The results for years 3–5 for the more advanced firms are positive although not statistically significant. For the less advanced firms in years 3–5, S/E is clearly negative. For the more advanced firms, S/E is significantly better than for the less advanced firms for years 1–5, with significance at the 10% level for years 1 and 5 and at the 5% level for years 2–4. For both the full sample and the more advanced firms, there appears to be an overall improving trend during the postevent period. The median preevent and postevent averages are negative for both the less and more advanced firms and thus for the full sample. For the most advanced firms, the median difference in the preevent and postevent averages is positive but not significant. The median postevent excess average performance is greater for the more advanced firms than for the less advanced firms ($p = .04$).

Table 7 shows the excess percent change in sales, assets, and employees between years 0 and 4. The table shows that, in comparison with the control firms, the event firms grew for all three measures. This was also the case for both the less advanced and more advanced firms. All of the results are significant with strong significance ($p = .00$) for the full sample. For the more advanced firms, the percent growth in sales was significantly better than for the less advanced firms at the 5% level. It is interesting that the percent growth in the number of employees was smaller for the more advanced firms than for the less advanced firms.

Table 8 gives the results for excess actual performance for total inventory to sales (I/S) and total inventory to cost of goods sold (I/CGS).[13] For the I/S variable, the median excess actual performance for the full sample is negative during the postevent period, indicating lower inventory levels for the sample firms than for the controls (an

13. Compustat's cost of goods sold (CGS) often includes selling, general, and administrative expenses (SGA). When this is the case, the corresponding Compustat SGA variable is missing. To correct for this inconsistency, the variable used here (which we refer to as CGS) is actually the sum of Compustat's CGS and SGA variables. Also note that, for the I/S variable, the five firms with missing data for all years are service firms. For the I/CGS variable, the 14 firms with missing data for all years are service firms—only one service firm remains. Thus, the results in table 8 are very similar to the results for the manufacturing firms alone (table 17).

TABLE 5 Excess Performance for Actual Net Income per Employee and Operating Income per Employee

Variable and Event Year	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
	Median ($1,000/Employee)	*p*-sgn	*N*	Median ($1,000/Employee)	*p*-sgn	*N*	Median ($1,000/Employee)	*p*-sgn	*N*	*p*-wrs
NI/E:										
−5	−.48	.82	98	−.50	.79	57	−.47	.73	41	.50
−4	−.41	.86	100	−.35	.70	57	−.72	.89	43	.73
−3	−.77	.95	102	−.50	.70	59	−.92	.98	43	.68
−2	−.02	.58	105	.02	.45	62	−.15	.73	43	.63
−1	−1.16	.99	108	−.53	.92	64	−1.54	.97	44	.64
0	−1.20	.95	108	−1.56	.97	64	−1.16	.67	44	.13
1	−.83	.93	108	−1.02	.95	64	−.47	.67	44	.04
2	−.46	.78	105	−1.88	.85	61	.16	.56	44	.04
3	−.17	.54	90	−1.39	.72	47	.13	.38	43	.33
4	−.27	.59	71	−1.28	.76	33	.29	.44	38	.34
5	.70	.22	61	−.70	.65	27	.84	.11	34	.35
−5 to −1	−.72	.86	108	−.68	.73	64	−.78	.85	44	.66
1 to 5	−.30	.68	108	−1.43	.87	64	.47	.33	44	.03
Difference	.29	.19	108	−.04	.55	64	.88	.09	44	.03

OI/E:										
−5	.44	.22	84	.24	.39	51	.83	.24	33	.28
−4	−.23	.67	85	−.16	.61	51	−.78	.70	34	.43
−3	−.07	.58	87	.31	.50	53	−.70	.70	34	.64
−2	−.44	.62	90	−1.14	.66	56	−.07	.57	34	.38
−1	.22	.42	93	.26	.55	58	.22	.37	35	.31
0	−.26	.62	94	−.67	.74	58	.74	.43	36	.19
1	−.18	.62	94	−1.33	.82	58	2.68	.31	36	.08
2	1.39	.17	92	−.24	.55	56	2.68	.07	36	.06
3	2.78	.02	78	2.03	.18	43	3.63	.02	35	.18
4	2.04	.22	61	−2.86	.77	29	2.52	.06	32	.09
5	3.03	.05	51	1.27	.27	24	4.83	.06	27	.25
−5 to −1	.07	.34	93	.16	.45	58	.06	.37	35	.27
1 to 5	1.19	.09	94	−1.28	.74	58	3.68	.00	36	.01
Difference	.46	.34	93	−.22	.74	58	2.63	.09	35	.03

SOURCE.—The data used to construct the performance measures were obtained from Compustat.

NOTE.—NI/E = net income per employee and OI/E = operating income per employee. The rows labeled ''−5 to −1,'' ''1 to 5,'' and ''Difference'' give the results for the average preevent period performance, the average postevent period performance, and the difference of the preevent period and postevent period averages, respectively. Results are reported for the full example of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p*-sgn contain *p*-values for the one-sided sign test of the null hypothesis H_0: true median ≤ 0 against the alternative H_A: true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less advanced firms.

TABLE 6 Excess Performance for Actual Sales per Employee

Variable and Event Year	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
	Median ($1,000/Employee)	*p*-sgn	*N*	Median ($1,000/Employee)	*p*-sgn	*N*	Median ($1,000/Employee)	*p*-sgn	*N*	*p*-wrs
S/E:										
−5	−4.36	.96	98	−4.63	.97	57	−4.09	.73	41	.14
−4	−7.37	.97	100	−8.71	.99	57	−3.50	.62	43	.09
−3	−7.38	.99	102	−8.38	.99	59	−7.02	.89	43	.16
−2	−5.03	.99	105	−5.64	.99	62	−1.50	.82	43	.26
−1	−8.07	1.00	108	−8.07	1.00	64	−6.59	.85	44	.17
0	−6.80	.99	108	−9.56	.99	64	−3.30	.85	44	.10
1	−11.01	.98	108	−16.96	.99	64	−1.63	.67	44	.07
2	−5.74	.91	105	−8.70	.98	61	3.84	.44	44	.04
3	−5.90	.77	90	−15.68	.93	47	2.15	.38	43	.03
4	−1.89	.68	71	−15.45	.98	33	4.06	.13	38	.03
5	−3.33	.70	61	−12.18	.88	27	2.63	.43	34	.07
−5 to −1	−6.23	.99	108	−7.43	.99	64	−3.04	.85	44	.14
1 to 5	−5.60	.95	108	−8.59	.98	64	−.37	.56	44	.04
Difference	−2.43	.81	108	−4.84	.92	64	1.10	.44	44	.12

SOURCE.—The data used to construct the performance measures were obtained from Compustat.

NOTE.—S/E = sales per employee. The rows labeled ''−5 to −1,'' ''1 to 5,'' and ''Difference'' give the results for the average preevent period performance, the average postevent period performance, and the difference of the preevent period and postevent period averages, respectively. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p*-sgn contain *p*-values for the one-sided sign test of the null hypothesis H_0:true median ≤ 0 against the alternative H_A:true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less advanced firms.

TABLE 7 Excess Percent Change in Sales, Assets, and Employees for Years 0–4

	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
Variable	Median (%)	*p*-sgn	*N*	Median (%)	*p*-sgn	*N*	Median (%)	*p*-sgn	*N*	*p*-wrs
%ΔS	10.39	.00	72	5.54	.08	33	13.92	.01	39	.04
%ΔA	13.25	.00	72	12.98	.04	33	15.61	.00	39	.35
%ΔE	8.61	.00	71	9.23	.02	33	6.12	.07	38	.84

SOURCE.—The data used to construct the performance measures were obtained from Compustat.

NOTE.—%ΔS = excess percent change in sales, %ΔA = excess percent change in assets, and %ΔE = excess percent change in number of employees. Results are reported for the full example of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p*-sgn contain the *p*-values for the one-sided sign test of the null hypothesis H_0: true median ≤ 0 against the alternative H_A: true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less advanced firms.

improvement). The results are significant at the 5% level for years 0, 2, and 3. The more advanced firms show a similar but stronger pattern, with significantly reduced I/S levels for years 0–4. For the less advanced firms, there is no clear pattern. The excess I/S levels for the more advanced firms are significantly lower than for the less advanced firms for all postevent years except year 5. The median excess postevent average I/S levels are also significantly negative for the full sample ($p = .06$) and strongly significantly negative for the more advanced firms ($p = .00$). The excess postevent average inventory level is also significantly lower for the more advanced firms than for the less advanced firms ($p = .02$). The excess I/S levels for both the full sample and the more advanced firms are also negative during the preevent period, and there is some suggestion that excess I/S levels increase during the preevent period and subsequently decrease during the postevent period. The median differences in the preevent and postevent averages, however, are positive, although not significant. For I/CGS for the full sample, there is a similar pattern.

In summary, the results for net income, operating income, and sales per employee indicate improved performance in comparison with the controls for the more advanced firms in the postevent period. The median performance throughout the postevent period for all of these variables is greater for the more advanced firms than for the less advanced firms with a significant difference for excess S/E. Excess inventory is lower for the event firms than for the controls during the postevent period for both of the inventory variables examined. The inventory results are stronger for the more advanced firms.

TABLE 8 Excess Performance for Actual Total Inventory to Sales and Total Inventory to Cost of Goods Sold

	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
Variable and Event Year	Median (%)	*p*-sgn	*N*	Median (%)	*p*-sgn	*N*	Median (%)	*p*-sgn	*N*	*p*-wrs
I/S:										
−5	−1.02	.11	95	.47	.66	54	−1.77	.01	41	.07
−4	−.94	.03	95	−.38	.34	54	−2.86	.01	41	.10
−3	−.95	.06	98	.22	.60	57	−2.36	.01	41	.02
−2	−.76	.12	101	−.38	.45	60	−1.79	.06	41	.11
−1	−.81	.08	103	−.71	.22	61	−1.19	.14	42	.06
0	−.64	.04	103	−.11	.30	61	−1.43	.02	42	.03
1	−.52	.22	103	.86	.70	61	−1.95	.04	42	.01
2	−.88	.04	100	−.06	.45	58	−1.51	.01	42	.05
3	−1.58	.04	85	−.11	.56	44	−1.84	.01	41	.04
4	−1.18	.20	68	1.29	.86	31	−2.13	.02	37	.07
5	−.06	.45	58	.33	.79	25	−.60	.24	33	.13
−5 to −1	−1.04	.08	103	−.03	.50	61	−2.90	.02	42	.02
1 to 5	−.93	.06	103	.29	.60	61	−1.59	.00	42	.02
Difference	.03	.58	103	−.12	.40	61	.27	.78	42	.56
I/CGS:										
−5	−1.29	.04	85	−.66	.39	51	−3.66	.01	34	.07
−4	−1.33	.04	85	−.27	.50	51	−3.65	.01	34	.05
−3	−.90	.05	88	−.06	.55	54	−6.07	.01	34	.00
−2	−.43	.34	91	.95	.70	57	−3.20	.11	34	.03
−1	−1.43	.03	93	−1.05	.18	58	−4.74	.05	35	.01
0	−1.19	.04	94	−.22	.35	58	−4.13	.02	36	.01
1	−.94	.13	94	.69	.65	58	−3.72	.02	36	.01
2	−1.03	.17	92	.02	.55	56	−1.70	.07	36	.06
3	−1.23	.21	78	.83	.73	43	−2.77	.05	35	.04
4	−.12	.50	61	1.29	.93	29	−1.92	.11	32	.10
5	.01	.71	51	.38	.73	24	.01	.65	27	.23
−5 to −1	−1.43	.11	93	.24	.65	58	−4.74	.01	35	.01
1 to 5	−.87	.13	94	−.48	.45	58	−2.28	.07	36	.03
Difference	.19	.66	93	−.09	.45	58	.43	.84	35	.68

SOURCE.—The data used to construct the performance measures were obtained from Compustat.

NOTE.—I/S = total inventory to sales and I/CGS = total inventory to cost of goods sold. The rows labeled ''−5 to −1,'' ''1 to 5,'' and ''Difference'' give the results for the average preevent period performance, the average postevent period performance, and the difference of the preevent period and postevent period averages, respectively. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. CGS = the sum of the Compustat ''cost of goods sold'' and ''selling, general, and administrative expense'' data items. Note that in the Compustat database, ''cost of goods sold'' is often not reported separately from ''selling, general, and administrative expense.'' The columns labeled *p*-sgn contain *p*-values for the one-sided sign test of the null hypothesis H_0: true median ≥ 0 against the alternative H_A: true median < 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically smaller than that of the less advanced firms.

D. Excess Cumulative Daily Stock Returns

Table 9 shows the results for the excess cumulative continuously compounded with-dividend daily stock returns. The cumulative returns begin in July of event year 0. For the full sample, the median excess cumulative return is 21.02% in year 5. The improvement is strongly significant ($p = .00$). For the more advanced firms, the median excess cumulative returns are 17.28%, 18.48%, and 22.11% for years 3, 4, and 5, respectively, with $p = .00$ for all 3 years. The excess cumulative returns for the more advanced firms are also positive and significant at the 5% level for years 1 and 2. The excess cumulative returns are not statistically significant for the less advanced firms for years 1–4. Year 5 performance, however, is positive and significant ($p = .10$). The differences between the excess cumulative returns for more and less advanced firms are significant at the 5% level for years 2–4 and at the 10% level for year 1. Thus, consistent with the accounting variables, the cumulative stock returns indicate improved long-term performance for the TQM firms, with stronger results for the more advanced TQM firms.

Table 9 also shows the median preevent and postevent period excess average monthly returns and the median difference of the preevent and postevent averages. For both the full sample and subsamples of less and more advanced firms, the excess average preevent monthly returns are negative. For both the full sample and the more advanced firms, the postevent excess average return is strongly positive ($p = .00$). The median difference in the preevent and postevent averages is strongly significantly positive for both the full sample and the more advanced firms ($p = .00$). For the less advanced firms, the preevent and postevent difference is positive and significant at the 10% level. The postevent excess average monthly returns and the postevent and preevent difference averages are significantly larger for the more advanced firms ($p = .06$ and $p = .09$, respectively).

Table 10 shows an analysis of the excess cumulative returns based on the assumption that the distribution of the stock returns is multivariate normal. This analysis corrects for any correlation between the excess returns for firms with the same event year 0. There is very little industry and event-year clustering in the sample, and thus it would appear unlikely that such correlation would have any impact on the results. Comparison of table 10 with table 9, however, allows direct assessment of any such impact. The two tables show that the results are essentially the same, with similar patterns in the levels of significance, verifying that correlation due to event-year clustering does not have a consequential impact on the results.

E. Downsizing

This section examines whether the observed positive performance for the TQM firms might be explained by downsizing that took place in

TABLE 9 Excess Cumulative Continuously Compounded With-Dividend Daily Stock Returns

Variable and Event Year	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
	Median (%)	*p*-sr	*N*	Median (%)	*p*-sr	*N*	Median (%)	*p*-sr	*N*	*p*-wrs
Excess cumulative returns:										
1	3.19	.05	107	1.55	.28	63	4.72	.02	44	.10
2	3.82	.18	106	−2.80	.66	63	9.41	.02	43	.05
3	3.91	.10	103	−1.72	.78	60	17.28	.00	43	.00
4	6.04	.02	87	−1.43	.67	46	18.48	.00	41	.00
5	21.02	.00	69	14.20	.10	32	22.11	.00	37	.20
Excess average monthly returns:										
−5 to −1	−.24	.98	106	−.30	.83	63	−.22	.98	43	.59
1 to 5	.25	.00	108	.13	.16	64	.36	.00	44	.06
Difference	.50	.00	106	.36	.07	63	.66	.00	43	.09

SOURCE.—The daily stock returns used to construct the performance measures were obtained from the database compiled by the Center for Research in Securities Prices at the University of Chicago.

NOTE.—The rows labeled ''−5 to−1,'' ''1 to 5,'' and ''Difference'' give the results for the average monthly returns for the preevent period, the postevent period, and the difference of the preevent period and postevent period average monthly returns, respectively. The monthly returns are calculated by cumulating the daily returns in the month. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p*-sr contain *p*-values for the one-sided Wilcoxon signed-rank test of the null hypothesis H_0: true median ≤ 0 against the alternative H_A: true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less advanced firms.

TABLE 10 Excess Cumulative Continuously Compounded With-Dividend Daily Stock Returns: Covariance-Based Analysis

Variable and Event Year	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
	Mean (%)	*p*-cv	*N*	Mean (%)	*p*-cv	*N*	Mean (%)	*p*-cv	*N*	*p*-2
Excess cumulative returns:										
1	3.37	.17	97	1.17	.40	56	6.36	.11	41	.23
2	4.48	.17	97	−.71	.55	57	11.88	.04	40	.09
3	7.26	.11	99	−3.79	.67	57	22.26	.00	42	.01
4	12.91	.04	86	−1.83	.57	46	29.87	.00	40	.01
5	24.21	.01	69	15.71	.15	32	31.55	.00	37	.20

SOURCE.—The daily stock returns used to construct the performance measures were obtained from the database compiled by the Center for Research in Securities Prices at the University of Chicago.

NOTE.—Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p*-cv contain *p*-values for the one-sided test of the null hypothesis $H_0: \mu \leq 0$ against the alternative $H_A: \mu > 0$. The estimates of the standard deviations of the means used in the test statistics are based on estimates of the variance-covariance matrices for events in the same year. The variance-covariance estimates are calculated from 5 years of monthly returns, where the monthly returns are calculated by cumulating the daily returns in the month. The column labeled *p*-2 contains *p*-values for the one-sided two-sample test of the null hypothesis $H_0: \mu_{more} \leq \mu_{less}$ against the alternative $H_A: \mu_{more} > \mu_{less}$, where μ_{less} and μ_{more} are the true means of the less advanced and more advanced firms, respectively.

conjunction with or during the same period as the deployment of TQM. There appears to be little consensus concerning a uniform definition of the term ''downsizing.'' It is also debatable the extent to which some downsizing-like activity is a natural consequence of the development of quality management systems. Empowerment of employees, one of the principles of TQM, is likely to lead to elimination of some levels of management and supervision over time. The effects of downsizing are examined here by examining the relationship between performance and percent changes in the number of employees. We believe that this captures what is commonly meant by downsizing—major reductions in the number of employees.

In order for the data to support the hypothesis that downsizing drives the results, (1) a large number of the event firms should be downsizing in comparison with their control firms, (2) the firms that do not downsize should not show significant positive performance, and (3) the firms that do downsize in comparison with their controls should show significant positive performance consistent with downsizing driving the results, especially when compared with the firms that do not downsize.

Requirement 1 is not supported by the data. Overall, both the event firms and the control firms grow in terms of the number of employees during the postevent period. The median percent change in employees between years 0 and 4 is 5.4% for the event firms and 1.0% for the controls. For the more advanced events only, the 4-year percent growth in employees is 2.1% for the event firms and 1.0% for the controls. For the less advanced events, the percent growth in employees is 9.2%, while the growth in the control portfolios is again 1.0%. Thus, the number of employees grows for most firms in the sample, and most event firms grow faster than their control portfolios. This is consistent with the results for excess percent change in employees given in table 7.

Requirements 2 and 3 above are examined by comparing the performance of events with negative excess percent change in employees with those with positive excess percent change in employees. Tables 11 and 12 give the results for the accounting variables and stock returns, respectively. Examination of these tables shows that the required patterns are not evident. Specifically, requirement 2 is not supported by the data. The vast majority of the performance measures for the event firms that do not downsize are positive and frequently significant. In particular, the results for the excess cumulative stock returns in table 12 show significant (at the 5% level) positive excess returns for years 4 and 5 for both the full sample of firms that did not downsize and the more advanced firms that did not downsize. Thus, there is evidence of a positive association between the implementation of TQM and performance for firms that do not downsize.

Finally, requirement 3 is also not supported by the data. There is no clear pattern of improved performance for the firms with negative ex-

cess percent change in employees over those with positive excess percent change in employees, and none of the corresponding *p*-values comparing the ''negative'' and ''positive'' firms are significant. For the accounting variables for the full sample shown in table 11, a positive difference is observed for only two of the eight performance variables. It is interesting to note that for both the accounting measures and the stock returns for the less advanced subsample the performance for events with a negative excess percent change in employees is worse (except for S/E) than for those with a positive excess percent change in employees. For the more advanced events, the opposite tends to be the case.

F. Results for the Manufacturing Firms Only

The analysis described above was repeated omitting the 15 event firms (indicated in app. A, table A1) corresponding to predominantly service companies. Tables 13–21 present the results. These results are not specifically discussed here. Overall, however, the results for the manufacturing firms alone are stronger and have increased statistical significance.

G. Other Variables and Research Issues Examined

1. *Analysts' forecasts.* In order to provide some empirical validation for the use of Value Line analysts' forecasts, the mean-squared errors (MSEs) of the analysts' forecasts were compared with the MSEs of forecasts made by simple autoregression models of order 1 (AR 1 models). The AR 1 models were estimated for each firm using the time series of annual values obtained from Compustat for the 11 years prior to and ending with event year 0 (i.e., from $t = -10$ to $t = 0$). Forecasts were then made for postevent years 1–5. For years 1 and 2, the MSE of the forecast is the average squared difference between the forecast and the realized value. In order to parallel the long-term forecasts provided by the Value Line analysts, the AR 1 forecasts for years 3–5 are averaged and compared with the average value of the realized values for years 3–5 by computing the average of the squared differences. The MSEs for the Value Line analysts' forecasts were computed in a similar manner using data and forecasts obtained from Value Line. Missing values in the Compustat preevent data caused the loss of a number of firms in the calculation of the AR 1 MSEs. For comparability, such missing firms were also deleted from the calculation of the Value Line MSEs.

Table 22 shows the efficiencies of the AR 1 forecasts relative to the analysts' forecasts. The efficiency is defined as the ratio of the MSE of the analysts' forecasts to the MSE of the AR 1 forecasts. Efficiencies of less than 100% show superior performance of the Value Line forecasts. The analysis was conducted separately for the control firms and

TABLE 11 **Long-Term Performance for the Accounting Variables for TQM Firms with Positive and Negative Excess Percent Change in Employees**

Variable and Excess %ΔE	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
	Median	*p*	*N*	Median	*p*	*N*	Median	*p*	*N*	*p*-wrs
NI/S:										
−	.76	.29	30	−.26	.60	14	1.75	.23	16	.16
+	.57	.03	41	.71	.01	19	.40	.42	22	.68
		.59*			.85*			.33*		
NI/A:										
−	.63	.43	30	−.98	.79	14	2.51	.23	16	.02
+	1.65	.01	41	.53	.08	19	1.86	.07	22	.31
		.79*			.96*			.36*		
OI/S:										
−	.95	.29	30	−1.01	.60	14	1.86	.23	16	.21
+	−.02	.62	41	.08	.50	19	−.15	.74	22	.40
		.54*			.74*			.42*		
OI/A:										
−	2.28	.29	30	−2.03	.79	14	3.18	.11	16	.03
+	2.50	.01	41	.23	.32	19	3.12	.01	22	.10
		.88*			.93*			.70*		

S/A:										
−	−2.21	.71	30	−22.49	.91	14	6.33	.40	16	.01
+	4.57	.06	41	−.04	.68	19	9.95	.01	22	.03
		.95*			.93*			.80*		
NI/E:										
−	−.28	.71	30	−1.53	.91	14	.29	.40	16	.15
+	.79	.50	41	.87	.50	19	−.07	.58	22	.54
		.71*			.87*			.37*		
OI/E:										
−	1.51	.42	26	−5.73	.97	11	5.05	.06	15	.09
+	2.04	.25	35	1.99	.41	18	2.04	.31	17	.21
		.76*			.81*			.48*		
S/E:										
−	−2.58	.71	30	−11.27	.91	14	.13	.40	16	.29
+	−1.78	.62	41	−17.62	.97	19	6.26	.14	22	.04
		.45*			.34*			.70*		

SOURCE.—For the variables NI/S, NI/A, OI/S, OI/A, and S/A, the data and forecasts used to construct the performance measures were obtained from the Value Line Investment Survey. For the variables NI/E, OI/E, and S/E, the data used to construct the performance measures were obtained from Compustat.

NOTE.—For net income to sales (NI/S), net income to assets (NI/A), operating income to sales (OI/S), operating income to assets (OI/A), and sales to assets (S/A), the median excess unexpected performance (%) for postevent years 3–5 is reported. For net income per employee (NI/E), operating income per employee (OI/E), and sales per employee (S/E), the median excess actual performance for postevent year 4 is reported (in units of $1,000 per employee). The rows labeled ''−'' and ''+'' correspond to events with negative and positive excess percent change in employees (%ΔE) for years 0–4, respectively. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample 44 event firms with more advanced TQM systems. The columns labeled *p* contain *p*-values for the one-sided sign test of the null hypothesis H_0: true median ≤ 0 against the alternative H_A: true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less advanced firms.

* *p*-values are for Wilcoxon rank-sum tests of the null hypothesis that the underlying distribution for the events with negative excess %ΔE is not stochastically larger than that for the events with positive excess %ΔE.

TABLE 12 Excess Cumulative Continuously Compounded With-Dividend Daily Stock Returns for TQM Firms with Positive and Negative Excess Percent Change in Employees

Event Year and Excess %ΔE	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
	Median (%)	*p*	*N*	Median (%)	*p*	*N*	Median (%)	*p*	*N*	*p*-wrs
1:										
−	−.68	.33	30	−3.03	.72	14	9.75	.10	16	.13
+	1.90	.12	40	−.34	.32	18	4.12	.13	22	.31
		.69*			.86*			.36*		
2:										
−	−2.49	.75	29	−12.89	.96	14	4.01	.18	15	.03
+	2.41	.37	40	−5.93	.72	18	7.73	.16	22	.20
		.84*			.87*			.56*		
3:										
−	4.83	.52	29	−15.02	.98	14	17.28	.04	15	.01
+	3.78	.06	40	−2.92	.41	18	17.71	.07	22	.09
		.84*			.98*			.36*		
4:										
−	4.55	.18	29	−9.27	.92	14	34.86	.01	15	.00
+	8.00	.02	40	.88	.24	18	12.07	.03	22	.21
		.68*			.95*			.21*		
5:										
−	21.27	.05	29	7.74	.51	14	34.29	.01	15	.10
+	20.78	.01	40	24.03	.04	18	17.95	.04	22	.59
		.56*			.86*			.17*		

SOURCE.—The daily stock returns used to construct the performance measures were obtained from the database compiled by the Center for Research in Securities Prices at the University of Chicago.

NOTE.—The rows labeled ''−'' and ''+'' correspond to events with negative and positive excess percent change in employees (%ΔE) for years 0–4, respectively. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p* contain *p*-values for the one-sided Wilcoxon signed-rank test of the null hypothesis H_0: true median ≤ 0 against the alternative H_A: true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less advanced firms.

**p*-values for Wilcoxon rank-sum tests of the null hypothesis that the underlying distribution for the events with negative %ΔE is not stochastically larger than that for the events with positive %ΔE.

for the event firms because the MSEs for the event firms contain a bias component due to their subsequent implementation of TQM. The results for the control firms do not include this bias component. Because the accounting data contain outliers and other deviations from normality and MSEs are very sensitive to these problems, a ''robust efficiency'' was also computed after trimming the largest 5% of the squared deviations.

Table 22 shows a clear superiority of the analysts' forecasts. For the usual efficiency (based on the standard MSEs), there is no instance of superior performance of the AR 1 forecasts. For the robust efficiencies, there are only three instances (out of 30) where the AR 1 forecasts had

TABLE 13 Excess Unexpected Performance for the Accounting Variables: Manufacturing Firms Only

	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
Variable and Event Year	Median (%)	*p*-sgn	*N*	Median (%)	*p*-sgn	*N*	Median (%)	*p*-sgn	*N*	*p*-wrs
NI/S:										
1	.14	.27	93	−.07	.65	58	.53	.09	35	.08
2	.29	.15	93	.08	.45	58	.97	.09	35	.06
3–5	.72	.01	87	.64	.06	52	1.23	.05	35	.10
NI/A:										
1	.35	.20	93	.03	.55	58	.94	.09	35	.05
2	.51	.01	93	.40	.12	58	.59	.02	35	.09
3–5	1.65	.01	87	.52	.11	52	2.26	.01	35	.02
OI/S:										
1	−.13	.58	93	−.50	.96	58	.70	.05	35	.14
2	−.27	.89	93	−.48	.96	58	.07	.50	35	.12
3–5	0.33	.26	87	.16	.44	52	1.66	.25	35	.11
OI/A:										
1	.91	.11	93	−.34	.65	58	1.29	.01	35	.08
2	.03	.50	93	−.01	.55	58	.03	.50	35	.49
3–5	2.22	.00	87	.62	.11	52	3.28	.00	35	.01
S/A:										
1	1.15	.15	93	.46	.45	58	6.35	.09	35	.13
2	1.80	.15	93	.05	.55	58	5.35	.05	35	.19
3–5	4.57	.10	87	−.62	.66	52	8.54	.01	35	.01

SOURCE.—The data and forecasts used to construct the performance measures were obtained from the Value Line Investment Survey.

NOTE.—NI/S = net income to sales, NI/A = net income to assets, OI/S = operating income to sales, OI/A = operating income to assets, and S/A = sales to assets. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p*-sgn contain the *p*-values for the one-sided sign test of the null hypothesis H_0: true median ≤ 0 against the alternative H_A: true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less mature firms.

superior performance. Most of the robust efficiencies are well below 70%.

2. *Firm size and calendar year.* Possible effects of firm size and year of TQM implementation were also examined. There were no clear differences in performance between event firms in the lower half of the size distribution and those in the upper half or between events that occurred in 1987 or earlier and those that occurred after 1987.

3. *Quality awards.* To determine if the results might be biased by selection of potential firms for the sample based on quality-award-related search criteria, the analysis was also rerun deleting all firms that were identified as a result of quality awards. The results are essentially unchanged, showing the same patterns of significance as for the full sample. In addition, a separate analysis was performed for the 39 events collected as a part of the first phase (Jarrell and Easton 1997) of the

TABLE 14 Excess Performance for Actual Net Income per Employee and Operating Income per Employee: Manufacturing Firms Only

Variable and Event Year	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
	Median ($1,000/Employee)	*p*-sgn	*N*	Median ($1,000/Employee)	*p*-sgn	*N*	Median ($1,000/Employee)	*p*-sgn	*N*	*p*-wrs
NI/E:										
−5	.28	.37	84	.13	.50	51	.36	.36	33	.42
−4	−.25	.59	85	.42	.50	51	−.41	.70	34	.65
−3	−.44	.80	87	.05	.50	53	−.58	.94	34	.65
−2	.19	.23	90	.03	.34	56	.29	.30	34	.47
−1	−.49	.89	93	−.14	.74	58	−1.35	.91	35	.55
0	−1.20	.89	93	−1.24	.88	58	−1.19	.75	35	.11
1	−.82	.85	93	−.95	.88	58	−.40	.63	35	.06
2	.07	.50	91	−1.66	.66	56	.41	.37	35	.06
3	.29	.37	78	−.23	.56	44	.40	.30	34	.36
4	.79	.40	63	−1.28	.64	31	1.44	.30	32	.26
5	1.85	.11	54	.70	.50	25	2.74	.07	29	.41
−5 to −1	.03	.50	93	.17	.45	58	−.19	.63	35	.57
1 to 5	.50	.34	93	−1.26	.74	58	1.31	.09	35	.03
Difference	.27	.27	93	.07	.45	58	.71	.25	35	.06

OI/E:										
−5	.44	.22	84	.24	.39	51	.83	.24	33	.28
−4	−.23	.67	85	−.16	.61	51	−.78	.70	34	.43
−3	−.07	.58	87	.31	.50	53	−.70	.70	34	.64
−2	−.44	.62	90	−1.14	.66	56	−.07	.57	34	.38
−1	.22	.42	93	.26	.55	58	.22	.37	35	.31
0	−.13	.58	93	−.67	.74	58	.89	.37	35	.14
1	−.05	.58	93	−1.33	.82	58	3.38	.25	35	.05
2	1.41	.15	91	−.24	.55	56	3.02	.05	35	.04
3	2.80	.01	77	2.03	.18	43	3.94	.01	34	.13
4	2.19	.18	60	−2.86	.77	29	2.65	.04	31	.05
5	3.03	.05	51	1.27	.27	24	4.83	.06	27	.25
−5 to −1	.07	.34	93	.16	.45	58	.06	.37	35	.27
1 to 5	1.25	.07	93	−1.28	.74	58	3.79	.00	35	.00
Difference	.46	.34	93	−.22	.74	58	2.63	.09	35	.03

SOURCE.—The data used to construct the performance measures were obtained from Compustat.

NOTE.—NI/E = net income per employee and OI/E = operating income per employee. The rows labeled ''−5 to −1,'' ''1 to 5,'' and ''Difference'' give the results for the average preevent period performance, the average postevent period performance, and the difference of the preevent period and postevent period averages, respectively. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p*-sgn contain *p*-values for the one-sided sign test of the null hypothesis H_0: true median ≤ 0 against the alternative H_A: true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less advanced firms.

TABLE 15 Excess Performance for Actual Sales per Employee: Manufacturing Firms Only

Variable and Event Year	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
	Median ($1,000/Employee)	*p*-sgn	*N*	Median ($1,000/Employee)	*p*-sgn	*N*	Median ($1,000/Employee)	*p*-sgn	*N*	*p*-wrs
S/E:										
−5	−4.07	.88	84	−4.06	.92	51	−4.09	.64	33	.17
−4	−6.76	.96	85	−8.71	.98	51	−3.53	.70	34	.08
−3	−7.36	.99	87	−8.38	.99	53	−7.10	.89	34	.15
−2	−6.69	1.00	90	−6.69	.99	56	−5.04	.89	34	.21
−1	−8.18	1.00	93	−10.13	1.00	58	−4.86	.84	35	.11
0	−6.38	.98	93	−10.82	.98	58	−2.29	.84	35	.05
1	−10.35	.95	93	−16.01	.98	58	−.87	.63	35	.04
2	−5.82	.90	91	−9.58	.96	56	1.71	.50	35	.03
3	−4.38	.63	78	−14.60	.85	44	2.63	.30	34	.03
4	−1.78	.60	63	−10.71	.96	31	4.06	.11	32	.03
5	−.44	.55	54	−8.75	.79	25	2.96	.36	29	.07
−5 to −1	−6.83	.99	93	−7.43	.99	58	−4.38	.91	35	.11
1 to 5	−2.70	.93	93	−6.61	.96	58	−.76	.63	35	.03
Difference	−.45	.73	93	−7.38	.93	58	2.73	.25	35	.03

SOURCE.—The data used to construct the performance measures were obtained from Compustat.

NOTE.—S/E = sales per employee. The rows labeled ''−5 to −1,'' ''1 to 5,'' and ''Difference'' give the results for the average preevent period performance, the average postevent period performance, and the difference of the preevent period and postevent period averages, respectively. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p*-sgn contain *p*-values for the one-sided sign test of the null hypothesis H_0: true median ≤ 0 against the alternative H_A: true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less advanced firms.

TABLE 16 Excess Percent Change in Sales, Assets, and Employees for Years 0–4: Manufacturing Firms Only

	Full Sample ($N = 69$)		Less Advanced TQM Firms ($N = 31$)		More Advanced TQM Firms ($N = 38$)		
Variable	Median (%)	*p*-sgn	Median (%)	*p*-sgn	Median (%)	*p*-sgn	*p*-wrs
%ΔS	7.40	.02	5.25	.24	11.37	.02	.10
%ΔA	7.53	.01	8.33	.08	7.22	.04	.53
%ΔE	7.44	.03	13.68	.04	1.53	.21	.80

SOURCE.—The data used to construct the performance measures were obtained from Compustat.

NOTE.—%ΔS = excess percent change in sales, %ΔA = excess percent change in assets, and %ΔE = excess percent change in number of employees. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p*-sgn contain *p*-values for the one-sided sign test of the null hypothesis H_0: true median ≤ 0 against the alternative H_A: true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less advanced firms.

study, which were not based on any searches relating to quality awards, and the 69 events collected in phase 2. The results of these analyses are also consistent. This stability lends further validity to the overall analysis and results.

4. *Less and more advanced firms intrasample validation.* The differences in performance between the more advanced and less advanced subsamples of TQM firms represent a very important intrasample validation of the overall research design. Because, however, the development of the firms' TQM systems occurs over a multiyear period, it is possible that the subsequent development of a firm's TQM system could be influenced by financial performance early in the postevent period. If this were the case, this might create a kind of ''TQM survivorship'' bias where firms that had positive early financial results would be more likely to continue the kinds of efforts necessary to develop an advanced TQM system. There are a number of reasons why this is not likely. This hypothesis assumes that managers expect early overall financial success from their TQM initiatives, that early success does not diminish the perceived need for major organizational change, and that early success drives development of an advanced system as defined by the Baldrige Award criteria rather than just a continuation of initial efforts.

This issue was also examined empirically by conditioning on zero excess average year 1 and year 2 stock returns and examining whether or not the difference in performance between the more advanced and less advanced firms persists in years 3 and 4. Two subsamples were constructed, one of the more advanced firms and the other of the less advanced firms, with zero median year 1 and year 2 excess average

TABLE 17 Excess Performance for Actual Total Inventory to Sales and Total Inventory to Cost of Goods Sold: Manufacturing Firms Only

	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
Variable and Event Year	Median (%)	*p*-sgn	*N*	Median (%)	*p*-sgn	*N*	Median (%)	*p*-sgn	*N*	*p*-wrs
I/S:										
−5	−1.02	.14	85	.57	.71	51	−3.66	.01	34	.04
−4	−1.25	.04	85	−.29	.39	51	−3.59	.01	34	.04
−3	−1.11	.05	88	.34	.66	54	−4.84	.00	34	.00
−2	−1.09	.10	91	−.30	.50	57	−3.10	.03	34	.04
−1	−1.11	.07	93	−.71	.26	58	−3.65	.09	35	.02
0	−.76	.05	93	−.07	.35	58	−3.76	.02	35	.01
1	−.52	.27	93	1.03	.74	58	−4.66	.05	35	.00
2	−.94	.07	91	−.06	.45	56	−2.93	.02	35	.02
3	−2.06	.07	78	−.11	.56	44	−2.85	.01	34	.02
4	−1.89	.22	63	1.29	.86	31	−2.80	.03	32	.04
5	−.03	.55	54	.33	.79	25	−.87	.36	29	.11
−5 to −1	−1.19	.07	93	.08	.55	58	−3.69	.01	35	.00
1 to 5	−.93	.11	93	.42	.65	58	−2.72	.01	35	.01
Difference	.24	.66	93	−.09	.45	58	1.40	.84	35	.63
I/CGS:										
−5	−1.29	.04	85	−.66	.39	51	−3.66	.01	34	.07
−4	−1.33	.04	85	−.27	.50	51	−3.65	.01	34	.05
−3	−.90	.05	88	−.06	.55	54	−6.07	.01	34	.00
−2	−.43	.34	91	.95	.70	57	−3.20	.11	34	.03
−1	−1.43	.03	93	−1.05	.18	58	−4.74	.05	35	.01
0	−1.27	.03	93	−.22	.35	58	−5.18	.01	35	.01
1	−.95	.11	93	.69	.65	58	−4.11	.01	35	.00
2	−1.03	.15	91	.02	.55	56	−1.91	.05	35	.05
3	−1.52	.18	77	.83	.73	43	−2.78	.03	34	.04
4	−.15	.45	60	1.29	.93	29	−1.96	.08	31	.08
5	.01	.71	51	.38	.73	24	.01	.65	27	.23
−5 to −1	−1.43	.11	93	.24	.65	58	−4.74	.01	35	.01
1 to 5	−.92	.11	93	−.48	.45	58	−2.50	.05	35	.02
Difference	.19	.66	93	−.09	.45	58	.43	.84	35	.68

SOURCE.—The data used to construct the performance measures were obtained from Compustat.

NOTE.—I/S = total inventory to sales and I/CGS = total inventory to cost of goods sold. The rows labeled ''−5 to −1,'' ''1 to 5,'' and ''Difference'' give the results for the average preevent period performance, the average postevent period performance, and the difference of the preevent period and postevent period averages, respectively. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. CGS = the sum of the Compustat ''cost of goods sold'' and ''selling, general, and administrative expense'' data items. Note that, in the Compustat database, ''cost of goods sold'' is often not reported separately from ''selling, general, and administrative expense.'' The columns labeled *p*-sgn contain *p*-values for the one-sided sign test of the null hypothesis H_0: true median ≥ 0 against the alternative H_A: true median < 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically smaller than that of the less advanced firms.

TABLE 18 Excess Cumulative Continuously Compounded With-Dividend Daily Stock Returns: Manufacturing Firms Only

	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
Variable and Event Year	Median (%)	*p*-sr	*N*	Median (%)	*p*-sr	*N*	Median (%)	*p*-sr	*N*	*p*-wrs
Excess cumulative returns:										
1	2.71	.11	92	1.55	.33	57	2.72	.06	35	.16
2	5.40	.17	91	−.17	.55	57	9.74	.04	34	.09
3	5.79	.05	89	−.70	.60	55	17.44	.00	34	.00
4	7.14	.01	75	−1.07	.59	43	33.31	.00	32	.00
5	23.51	.00	61	19.07	.11	30	31.21	.00	31	.10
Excess average monthly returns:										
−5 to −1	−.33	.98	91	−.30	.83	57	−.47	.99	34	.70
1 to 5	.34	.00	93	.17	.11	58	.52	.00	35	.05
Difference	.55	.00	91	.41	.04	57	.94	.00	34	.04

SOURCE.—The daily stock returns used to construct the performance measures were obtained from the database compiled by the Center for Research in Securities Prices at the University of Chicago.

NOTE.—The rows labeled ''−5 to −1,'' ''1 to 5,'' and ''Difference'' give the results for the average monthly returns for the preevent period, the postevent period, and the difference of the preevent period and postevent period average monthly returns, respectively. The monthly returns are calculated by cumulating the daily returns in the month. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p*-sr contain the *p*-values for the one-sided Wilcoxon signed-rank test of the null hypothesis H_0: true median ≤ 0 against the alternative H_A: true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less advanced firms.

TABLE 19 Excess Cumulative Continuously Compounded With-Dividend Daily Stock Returns: Covariance-Based Analysis for the Manufacturing Firms Only

	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
Variable and Event Year	Mean (%)	*p*-cv	*N*	Mean (%)	*p*-cv	*N*	Mean (%)	*p*-cv	*N*	*p*-2
Excess cumulative returns:										
1	2.87	.23	82	.76	.44	50	6.15	.16	32	.25
2	4.83	.18	82	.71	.46	51	11.59	.09	31	.16
3	9.90	.07	85	−.99	.54	52	27.04	.00	33	.02
4	15.67	.03	74	−.42	.52	43	37.99	.00	31	.01
5	28.64	.00	61	16.16	.15	30	40.72	.00	31	.12

SOURCE.—The daily stock returns used to construct the performance measures were obtained from the database compiled by the Center for Research in Securities Prices at the University of Chicago.

NOTE.—Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p*-cv contain *p*-values for the one-sided test of the null hypothesis H_0: $\mu \leq 0$ against the alternative H_A: $\mu > 0$. The estimates of the standard deviations of the means used in the test statistics are based on estimates of the variance-covariance matrices for events in the same year. The variance-covariance estimates are calculated from 5 years of monthly returns, where the monthly returns are calculated by cumulating the daily returns in the month. The column labeled *p*-2 contains *p*-values for the one-sided two-sample test of the null hypothesis H_0: $\mu_{more} \leq \mu_{less}$ against the alternative H_A: $\mu_{more} > \mu_{less}$, where μ_{less} and μ_{more} are the true means of the less advanced and more advanced firms, respectively.

stock returns. For these subsamples, the year 3 and year 4 performance of the more advanced firms continues to be significantly better than for the less advanced firms. This analysis provides evidence against the hypothesis that ''feedback'' due to early financial performance is the driver of the difference in the results for the more and less advanced firms.

VII. Conclusion

The major finding of this study is clear evidence that the long-term performance of firms that implemented TQM is improved. We believe the evidence of improvement is particulary strong when the overall analysis is considered. Specifically, both the results based on the excess unexpected performance of the accounting variables and on excess cumulative stock returns are consistent. We also view the overall stronger performance of the more advanced TQM firms, which were identified independently by interviews, as both an important test of the research methodology and compelling evidence that management methods that constitute TQM are associated with improved performance. In addition, the results are even stronger when the analysis is limited to just manufacturing firms. The study has also examined whether downsizing, which might have occurred in conjunction with the implementation of

TABLE 20 Long-Term Performance for the Accounting Variables for TQM Firms with Positive and Negative Excess Percent Change in Employees: Manufacturing Firms Only

Variable and Excess %ΔE	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
	Median	*p*	*N*	Median	*p*	*N*	Median	*p*	*N*	*p*-wrs
NI/S:										
−	1.12	.17	28	.32	.50	13	2.26	.15	15	.13
+	.62	.02	35	.71	.02	18	.55	.31	17	.60
		.55*			.81*			.34*		
NI/A:										
−	1.00	.29	28	−.30	.71	13	2.76	.15	15	.03
+	2.08	.01	35	1.23	.12	18	2.13	.02	17	.16
		.82*			.94*			.52*		
OI/S:										
−	1.19	.29	28	−2.39	.71	13	2.06	.15	15	.20
+	.08	.50	35	−.17	.59	18	.14	.50	17	.13
		.66*			.77*			.59*		
OI/A:										
−	2.37	.17	28	−3.14	.71	13	3.28	.06	15	.03
+	2.86	.02	35	.36	.41	18	6.33	.01	17	.04
		.91*			.94*			.80*		
S/A:										
−	1.74	.57	28	−26.74	.87	13	6.41	.30	15	.02
+	1.40	.16	35	−2.65	.76	18	9.07	.02	17	.04
		.90*			.92*			.71*		
NI/E:										
−	−.35	.71	28	−1.79	.87	13	.41	.50	15	.16
+	2.10	.25	35	2.16	.41	18	2.10	.31	17	.44
		.87*			.90*			.57*		
OI/E:										
−	1.51	.42	26	−5.73	.97	11	5.05	.06	15	.09
+	2.35	.20	34	1.99	.41	18	2.35	.23	16	.12
		.82*			.81*			.60*		
S/E:										
−	−.94	.57	28	−7.08	.87	13	.25	.30	15	.24
+	−1.78	.63	35	−14.16	.95	18	4.44	.17	17	.05
		.31*			.35*			.56*		

SOURCE.—For the variables NI/S, NI/A, OI/S, OI/A, and S/A, the data and forecasts used to construct the performance measures were obtained from the Value Line Investment Survey. For the variables NI/E, OI/E, and S/E, the data used to construct the performance measures were obtained from Compustat.

NOTE.—For net income to sales (NI/S), net income to assets (NI/A), operating income to sales (OI/S), operating income to assets (OI/A), and sales to assets (S/A), the median excess unexpected performance (%) for postevent years 3–5 is reported. For net income per employee (NI/E), operating income per employee (OI/E), and sales per employee (S/E), the median excess actual performance for postevent year 4 is reported (in units of $1,000 per employee). The rows labeled "−" and "+" correspond to events with negative and positive excess percent change in employees (%ΔE) for years 0–4, respectively. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p* contain *p*-values for the one-sided sign test of the null hypothesis H_0: true median ≤ 0 against the alternative H_A: true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less advanced firms.

**p*-values are for Wilcoxon rank-sum tests of the null hypothesis that the underlying distribution for the events with negative %ΔE is not stochastically larger than that for the events with positive %ΔE.

TABLE 21 Excess Cumulative Continuously Compounded With-Dividend Daily Stock Returns for TQM Firms with Positive and Negative Excess Percent Change in Employees: Manufacturing Firms Only

	Full Sample			Less Advanced TQM Firms			More Advanced TQM Firms			
Event Year and Excess %ΔE	Median (%)	*p*	*N*	Median (%)	*p*	*N*	Median (%)	*p*	*N*	*p*-wrs
1:										
−	−.79	.39	28	−3.13	.70	13	2.70	.15	15	.17
+	1.90	.14	34	1.07	.29	17	2.72	.17	17	.30
		.74*			.83*			.48*		
2:										
−	−7.24	.76	27	−15.64	.97	13	6.71	.17	14	.03
+	5.20	.33	34	−5.44	.71	17	10.06	.12	17	.17
		.86*			.89*			.62*		
3:										
−	4.83	.48	27	−17.15	.98	13	17.44	.02	14	.00
+	5.89	.05	34	−.70	.34	17	18.75	.05	17	.08
		.85*			.99*			.41*		
4:										
−	6.50	.17	27	−13.45	.93	13	36.83	.01	14	.00
+	10.66	.02	34	2.84	.22	17	18.14	.02	17	.12
		.72*			.96*			.31*		
5:										
−	23.59	.06	27	3.97	.54	13	41.07	.01	14	.09
+	22.81	.00	34	24.54	.05	17	21.02	.02	17	.37
		.69*			.87*			.35*		

SOURCE.—The daily stock returns used to construct the performance measures were obtained from the database compiled by the Center for Research in Securities Prices at the University of Chicago.

NOTE.—The rows labeled ''−'' and ''+'' correspond to events with negative and positive excess percent change in employees (%ΔE) for years 0–4, respectively. Results are reported for the full sample of 108 events, the subsample of 64 event firms with less advanced TQM systems, and the subsample of 44 event firms with more advanced TQM systems. The columns labeled *p* contain *p*-values for the one-sided Wilcoxon signed-rank test of the null hypothesis H_0: true median ≤ 0 against the alternative H_A: true median > 0. The column labeled *p*-wrs contains *p*-values for the one-sided Wilcoxon rank-sum test of the null hypothesis that the underlying distribution of the more advanced firms is not stochastically larger than that of the less advanced firms.

**p*-values for Wilcoxon rank-sum tests of the null hypothesis that the underlying distribution for the events with negative excess %ΔE is not stochastically larger than that for the events with positive excess %ΔE.

TQM, could explain the positive performance we observed. This hypothesis is not supported by the data.

While no observational study can prove a causal relationship, this study is based on a carefully developed research methodology designed to provide as compelling evidence as possible on the impact of the adoption of TQM on corporate financial performance. Specifically, a carefully controlled event-study approach is used rather than cross-sectional analysis; the sample of TQM firms is selected on the basis of in-depth interviews, not self-selected on the basis of mail survey responses or public pronouncements; an established operational definition of a TQM system is the basis for selection (the Baldrige Award

TABLE 22 Efficiencies of AR 1 Forecasts Relative to Value Line Analysts' Forecasts for the Accounting Variables

Variable and Event Year	Control Firms			Event Firms		
	Eff (%)	r-Eff (%)	*N*	Eff (%)	r-Eff (%)	*N*
NI/S:						
1	44.44	54.21	299	5.33	30.76	98
2	25.01	36.25	288	1.65	30.96	96
3–5	1.84	71.41	248	.00	41.18	84
NI/A:						
1	47.63	50.35	299	20.72	55.06	98
2	30.50	37.90	288	5.72	49.32	96
3–5	4.11	92.48	248	.09	82.70	84
OI/S:						
1	75.47	77.96	259	83.51	66.18	85
2	34.94	37.70	249	37.03	82.96	84
3–5	35.51	94.83	212	7.53	79.70	72
OI/A:						
1	72.61	82.46	259	96.25	68.19	85
2	43.09	52.28	249	63.07	114.90	84
3–5	61.11	188.40	212	13.58	110.44	72
S/A:						
1	50.57	66.11	299	79.37	75.39	98
2	65.71	50.31	288	51.24	53.90	96
3–5	83.12	95.02	248	40.07	75.41	84

SOURCE.—The performance measures and forecasts used to calculate the mean-squared errors for the analysts' forecasts were obtained from the Value Line Investment Survey. The performance measures used to calculate the AR 1 forecasts were obtained from Compustat.

NOTE.—Efficiencies (Eff) and ''robust'' efficiencies (r-Eff) are reported for AR 1 time-series forecasts for net income to sales (NI/S), net income to assets (NI/A), operating income to sales (OI/S), operating income to assets (OI/A), and sales to assets (S/A) for postevent years 1, 2, and the average of years 3–5. The efficiency is calculated as the mean-squared error of the Value Line analysts' forecasts divided by the mean-squared error of the time-series forecasts. The robust efficiencies are calculated in a similar fashion except that the mean-squared errors are replaced by 5% trimmed mean-squared errors (the mean-squared error obtained after omitting the largest 5% of the squared errors). The AR 1 model was estimated for each firm based on 11 years of data prior to and ending with event year 0 (i.e., t_{-10} to t_0).

criteria); and the approach is further validated by comparison of the more and less advanced TQM firms. In addition, there is a plausible causal mechanism for the observed improvement performance—TQM, after all, does focus specifically on generating quality and operational improvements. Further, the management changes associated with the development of a TQM system are sufficient in scope that it is plausible that their effects are observable in overall corporate performance. Finally, even under the most unfavorable interpretation, the results of this study clearly provide evidence against the proposition that implementation of TQM actually hurts corporate performance.

It is important, however, to recognize limitations on the generalizability of the results. This study examines whether TQM is associated with an improvement in financial performance for companies that made

serious efforts to implement TQM. This was done by comparing actual performance with a carefully constructed benchmark of what performance would have been without TQM. The finding that TQM improves performance for the companies that implement it, however, cannot necessarily be generalized to a prescription that the companies that did not implement TQM would also have improved performance if they had. It is possible that there are enabling factors that would make TQM effective in some companies and ineffective in others. The decision to implement or not implement TQM may be based on managers' knowledge of these factors.

Appendix A

TABLE A1 Sample of TQM Firms

TQM Firm	Year of Implementation
ADC Communications	1987
Advanced Micro Devices	1988
Air Products	1987
Albany International Corporation	1987
Alcoa	1990
Allied Signal	1991
Amdahl Corporation	1984
American Express*	1989
Analog Devices	1987
Applied Materials	1985
Arkansas Best Corporation*	1984
Armstrong World Industries	1983
Arvin Industries	1986
AT & T*	1988
Baldor Electric	1987
Banc One Corporation*	1986
Bausch & Lomb	1989
Baxter International	1985
Black & Decker	1990
Boise Cascade	1990
Cameron Iron Works	1984
Carolina Freight Corporation*	1984
Carpenter Technology Corporation	1987
Caterpillar	1983
Ceridian	1984
Chevron	1987
Chrysler	1985
Conner Peripherals	1989
Consolidated Freight*	1990
Corning Glass	1984
Cummins Engine	1983
Dana Corporation	1984
Diebold	1990
Digital Equipment	1989
Dun & Bradstreet*	1991
DuPont	1987
Eastman Kodak	1983

TABLE A1 (*Continued*)

TQM Firm	Year of Implementation
Ethyl Corporation	1986
Federal Express*	1986
Firestone Tire	1982
First Chicago*	1985
Fluke (John) Manufacturing	1990
Ford Motor	1984
FPL Group*	1986
Gaylord Container Corporation	1990
General Datacomm	1987
General Motors	1985
Goodyear Tire	1990
Goulds Pumps	1989
Grumman	1988
GTE Corporation*	1986
Hanna (M.A)	1990
Harris Corporation	1986
Hewlett Packard	1983
Hillenbrand Industries	1987
Hormel	1986
IBM	1989
Integrated Device Technology	1989
Intel	1985
International Paper	1985
James River	1986
Johnson Controls	1986
Kulicke and Soffa	1988
Lubrizol Corporation	1988
Lyondell Petroleum	1989
Micron Technology	1988
Millipore Corporation	1986
Minnesota Mining & Manufacturing (3M)	1984
Molex	1986
Moog	1989
Morton International	1991
Motorola	1983
Nashua Corporation	1981
National Semiconductor	1990
Pacific Telesis*	1989
Perkin Elmer Corporation	1984
PPG Industries	1986
Proctor & Gamble	1987
Raychem Corporation	1987
Roadway Services*	1989
Rockwell International Corporation	1986
Rogers Corporation	1983
Rohr Industries	1989
Scotsman Industries	1990
Sealed Air Corporation	1989
Snap-on Tools	1986
Square D	1987
Standard Register	1989
Sterling Chemical	1990
Storage Technology	1988
Sun Microsystems	1988
Tektronix	1989

TABLE A1 (*Continued*)

TQM Firm	Year of Implementation
Teradyne	1990
Texas Instruments	1982
Thomas & Betts	1987
Timken Company	1983
Union Camp Corporation	19871987
Union Carbide	1988
Unisys	1988
United Technologies	1984
Varian Associates	1987
VLSI Technology	1989
Westinghouse Electric Corporation	1982
Weyerhaeuser	1989
Whirlpool Corporation	1990
WPL Holdings*	1987
Xerox Corporation	1983
Yellow Corporation*	1990

* Indicates a predominantly service company.

Appendix B

TABLE B1 Interview Topics

General Category and Specific Approach
Training:
Senior management training
Awareness training
Training of other management levels
Workforce basic training
Technical training
Training for engineering
Teams:
Workforce improvement teams
Natural work-group teams
Cross-functional teams
Vertical teams
Work-cell teams
Self-managed teams
Project-oriented teams
Management teams
Customers:
Customer satisfaction surveys
Customer complaint tracking
Customer audits
Organizational structures:
Senior management quality council
Departmental quality councils
Specific-location quality councils
Internal quality consultants
Planning and values:
Written quality values and/or mission statement
Hoshin planning or policy deployment
Formal benchmarking
Quality and customer satisfaction measures reported to senior management
Audits:
Quality assurance audits
ISO 9000
Baldrige self-assessments
Other management systems audits
Team processes and tools:
Problem-solving process
Flowcharting
Plan-do-check-act
Seven basic quality control tools
Root-cause analysis

TABLE B1 (*Continued*)

General Category and Specific Approach
Involvement and morale:
Suggestion systems
Employee quality recognition
Employee morale survey
Design and engineering:
Design-for-manufacturability
Concurrent or simultaneous engineering
Design of experiments
Taguchi methods
Quality function deployment
Production:
SPC
JIT, cycle-time reduction, or single minute exchange of die
Activity-based costing
Work cells
Suppliers:
Supplier tracking
Supplier certification
Supplier quality audits
Supplier training
Joint supplier teams
Ship-to-stock, or ship-to-production relationships
Supplier integration into product development
Crosby:
Quality improvement teams
Error cause removal system
Corrective action teams
Cost of quality, or price of nonconformance
Measure and display
Quality education system training
Zero defects days

References

Anderson, S. W.; Daly, J.; and Johnson, M. F. 1995. The value of management control systems: Evidence on the market reaction to ISO 9000 quality assurance certification. Working paper. Ann Arbor: University of Michigan Business School.

Barber, B. M., and Lyon, J. D. 1996. Detecting abnormal operating performance: The empirical power and specification of test statistics. *Journal of Financial Economics* 41, no. 3:359–99.

Brown, L. D. 1993. Earnings forecasting research: Its implication for capital markets research. *International Journal of Forecasting* 9:295–320.

Brown, L. D.; Hagerman, R.; Griffin, P.; and Zmijewski, M. 1987. Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. *Journal of Accounting and Economics* 9:61–87.

Brown, L. D., and Rozeff, M. 1978. The superiority of analyst forecasts as measures of expectations: Evidence from earnings. *Journal of Finance* 6:1–16.

Easton, G. S. 1995. A Baldrige examiner's assessment of U.S. total quality management. In R. E. Cole (ed.), *The Death and Life of the American Quality Movement.* New York: Oxford University Press.

Easton, G. S., and Jarrell, S. L. Forthcoming. The emerging academic research on the link between total quality management and corporate financial performance: A critical review, in *Perspectives in Total Quality,* M. J. Stahl (ed.), Blackwell Publishers and Quality Press.

Fried, D., and Givoly, D. 1982. Financial analysts' forecasts of earnings: A better surrogate for market expectations. *Journal of Accounting and Economics* 4:85–108.

Healy, P.; Palepu, K.; and Ruback, R. 1992. Does corporate performance improve after mergers? *Journal of Financial Economics* 31:135–75.

Hendricks, K. B., and Singhal, V. R. 1996. Quality awards and the market value of the firm: An empirical investigation. *Management Science* 42, no. 3:415–36.

Jarrell, S. L. 1991. Do takeovers generate value? Non-stock price evidence on the ability of the capital market to assess takeovers. Ph.D. dissertation, University of Chicago, Graduate School of Business.

Jarrell, S. L., and Easton, G. S. 1997. An exploratory empirical investigation of the effects of total quality management on corporate performance. In P. Lederer and U. Karmarkar (eds.), *The Practice of Quality Management.* Norwell, Mass.: Kluwer.

National Institute of Standards and Technology (NIST). 1994. *Malcolm Baldrige National Quality Award 1995 Award Criteria.* Washington, D.C.: U.S. Department of Commerce, NIST.

Schipper, K. 1991. Commentary on analysts' forecasts. *Accounting Horizons* 5:105–21.

U.S. General Accounting Office. 1991. Management practices: U.S. companies improve performance through quality efforts. Report no. GAO/NSIAD-91-190. Washington, D.C.: U.S. General Accounting Office.