The Financial Impact of ISO 9000 Certification in the United States: An Empirical Analysis

Charles J. Corbett
Anderson School of Management, University of California, Los Angeles, California 90095-1481,
charles.corbett@anderson.ucla.edu

María J. Montes-Sancho
Universidad Carlos III de Madrid, Madrid 126–28903, Getafe, Madrid, Spain, mmontes@emp.uc3m.es

David A. Kirsch
R.H. Smith School of Business, University of Maryland, College Park, Maryland 20742, dkirsch@rhsmith.umd.edu

The ISO 9000 series of quality management systems standards, introduced in 1986, has been adopted at over 560,000 locations worldwide. Anecdotal evidence suggests that firms can achieve internal benefits such as quality or productivity improvements or that certification can help firms maintain or increase their market share, or both. Others argue that the standard is too generic to cause performance improvement but can be seen as a signal of good management. In this paper, we track financial performance from 1987 to 1997 of all publicly traded ISO 9000 certified manufacturing firms in the United States with SIC codes 2000–3999, and test whether ISO 9000 certification leads to productivity improvements, market benefits, and improved financial performance. We employ event-study methods, matching each certified firm to a control group of one or more noncertified firms in the same industry with similar precertification size and/or return on assets. We find that firms' decision to seek their first ISO 9000 certification was indeed followed by significant abnormal improvements in financial performance, though the exact timing and magnitude of this effect depend on the specification of the control group. Three years after certification, the certified firms do display strongly significant abnormal performance under all control-group specifications. The degree to which the precise results vary across control-group specifications indicates that event studies should always include extensive sensitivity analysis, for instance matching by size and performance separately and jointly, using both single firms and portfolios as controls.

Key words: ISO 9000; quality management; standards; financial; empirical; event study; Compustat
History: Accepted by William S. Lovejoy, operations and supply chain management; received June 21, 2002. This paper was with the authors 11 months for 3 revisions.

1. Introduction

The ISO 9000 series of quality management systems standards is an exceptionally widely diffused management practice. By December 2001, it had been adopted by over 560,000 facilities in 159 countries; over 38,000 of those facilities are located in the United States (ISO 2003). Despite this, considerable uncertainty exists among practitioners and scholars about the true effects of the standard, given its generic and minimally prescriptive nature. Reimann and Hertz (1994) explain ISO 9000's focus on conformance and how it differs from the Baldrige award's much broader focus on competitiveness. Moreover, implementing ISO 9000 is not without cost. Besides the fees for auditors and consultants, the process generally consumes considerable employee time and effort. Especially as ISO 9000 was the first standard of its type, many firms had to develop, implement, and document their quality management system from scratch.

Remarkably, though, there is no hard evidence to date on the financial effects of ISO 9000 certification. Landmark studies by Hendricks and Singhal (1997) and Easton and Jarrell (1998) show that total quality management (TQM) leads to improved financial performance, but the philosophy underlying TQM is quite different from (though not inconsistent with) that of ISO 9000. TQM is a loosely defined, holistic set of principles and tools aimed at, among others, encouraging continuous improvement and prevention of defects; the premise of ISO 9000 is that well defined and documented procedures improve consistency of output. Degree of TQM implementation is difficult to assess objectively: Hendricks and Singhal (1997) use quality awards to signal the presence of an effective TQM program, and Hendricks and Singhal (2001b) find that the exact nature of the quality award influences the extent to which recipient firms benefit from it. Easton and Jarrell (1998) use in-depth

questionnaires to assess degree of TQM implementation. ISO 9000 certification, on the other hand, is unambiguous, due to the third-party audits involved. Anecdotal evidence suggests that some firms truly take the underlying principles of the ISO 9000 standards to heart, while others only do the minimum necessary to pass the audits. Indeed, Andrews et al. (2001) find noticeable differences beween auditors' interpretations of the ISO 14001 standards, the environmental management systems counterpart to ISO 9000. Some of the areas of variation were specific to ISO 14001, while others are likely to apply equally to ISO 9000, such as variations in auditor training and experience. In this study, we focus only on the effects of certification, ignoring the depth of implementation.

Most scholarly work so far on the benefits of ISO 9000 has relied on surveys; certified firms often claim to have obtained benefits from adopting ISO 9000, but these benefits are self-reported. In this paper, we combine Compustat data with a database of all ISO 9000 certifications in the United States in the manufacturing sector, with SIC codes 2000–3999. The result is a database of 7,238 firms for the 1987–1997 period. We perform a range of sensitivity analyses to explore the impact of different matching methods. We match certified firms to individual noncertified firms and to control groups of noncertified firms. We use several criteria to select control firms: precertification size, return on assets, and both. We use various parametric and nonparametric test statistics. In all cases, we find that three years after their first certification, firms experienced strongly significant abnormal performance, though the exact timing and magnitude of the effects vary with different control-group specifications.

In §2, we review relevant literature related to ISO 9000 and TQM. Section 3 develops the theoretical relationships between ISO 9000 certification and financial performance. Section 4 describes our data. Section 5 describes the event-study methods used and presents our main results. We discuss these and our sensitivity analysis in §6. Section 7 offers our conclusions and suggestions for future research.

2. Literature

In this section we review literature on ISO 9000 certification and on TQM. Most literature on ISO 9000 either focuses on adoption or uses self-reported survey-based performance measures. Surveys by Terziovski et al. (1997) and Singels et al. (2001) find no link between ISO 9000 and organizational performance among Australian and Dutch firms, respectively, though Terziovski et al. (2003) do find that higher reported business performance is positively associated with a broader set of motivations for seeking ISO 9000 certification. Casadesús et al. (2001) find that

Basque companies report substantial operational and financial benefits from certification.

Using Compustat data, Anderson et al. (1999) find that U.S. firms with higher exports to Europe are more likely to seek ISO 9000 certification, and Corbett (2004), using data from a multinational survey, finds that exports drive early certifications; Guler et al. (2000), Mendel (2001), and Corbett and Kirsch (2001) examine drivers of global diffusion of ISO 9000 and ISO 14000. Adams (1999) finds that ISO 9000 certification in New Zealand increases with firm size, Tobin's Q, and market segmentation. Naveh and Marcus (2004), using a detailed survey of ISO 9000 in the United States, find that "going beyond" the requirements of the standard increases its value to the firm. Terlaak and King (2005a) argue that ISO 9000 is a signalling mechanism and finds sales growth that is partly consistent with that view; Terlaak and King (2005b) also find evidence that institutional theory contributes to explaining adoption. King and Lenox (2001) and Russo (2001) find some evidence that ISO 9000 and ISO 14000 certification respectively improve firms' environmental performance, suggesting that both standards do also have an impact on production practices.

Several authors have examined the financial effects of implementing TQM programs. Hendricks and Singhal (1996, 1997, 2001a) use event studies to show that winning a quality award increases short-term and long-term stock price and operating performance. Hendricks and Singhal (2001b) find that the extent of financial benefits arising from TQM depends on firm characteristics and on the type of quality award involved. The event studies in Easton and Jarrell (1998) also show a significant improvement in performance after implementing TQM. Powell (1995) concludes that the tacit resources often associated with successful TQM implementation improve performance, rather than the TQM tools themselves. Ittner et al. (2001) report that proactive investments in prevention do lead to lower nonconformance quality costs.

By contrast, very few authors have explicitly measured the financial effects of ISO 9000 certification. Event studies by Lima et al. (2000) and Martínez-Costa and Martínez-Lorente (2002) find no effects in Brazil and Spain, respectively, while the Spanish event study by Nicolau and Sellers (2002) only finds a positive stock price effect on the day of certification itself. Docking and Dowen (1999) find that small firms in the United States experienced positive stock market reaction to the announcement of their first ISO 9000 certification, but that larger firms' stock price did not respond. Simmons and White (1999) find that ISO 9000 certification is associated with superior financial performance in the United States within SIC

code 36 (electronic and other electrical equipment and components, except computer equipment), but have insufficient longitudinal data to indicate the direction of causality. Heras et al. (2002) find that more profitable firms in the Basque region are more likely to seek ISO 9000 certification, but that they do not become more profitable as a result. By using the much larger U.S. database, focusing on long-term operating performance, and matching firms based on precertification performance, we are able to detect significant and persistent effects of certification.

3. Theory Development and Hypotheses

Our theoretical basis is analogous to that commonly found in the TQM literature, for instance in Hendricks and Singhal (1997) and Easton and Jarrell (1998). The starting point is the hypothesis that ISO 9000 certification has a positive effect on profitability, regardless of the underlying mechanism, where profitability is commonly measured by return on assets (ROA: operating income/assets) or return on sales (ROS: operating income/sales), both before depreciation.

HYPOTHESIS 1. *ISO 9000 certification leads to an increase in return on assets (ROA) and in return on sales (ROS).*

A corollary of this hypothesis is that certification contributes to the stock market's valuation of a firm's intangible assets, measured by Tobin's Q; a firm with the same physical assets but "better management" should have a higher market valuation. Tobin's Q is defined as "the ratio of the market value of the outstanding financial claims on the firm to the current replacement cost of the firm's assets" (Lewellen and Badrinath 1997, p. 78), and measures the alternative-use value of the firm's assets. For the horizon we consider here, Tobin's Q can also be thought of as a proxy for stock price movement.

HYPOTHESIS 2. *ISO 9000 certification leads to an increase in Tobin's Q.*

Exactly how is ISO 9000 certification linked to a firm's financial performance? The academic and practitioner literatures propose several possible links. ISO 9000 may improve a firm's internal procedures, and hence enhance performance by reducing cost. Alternatively, ISO 9000 certification may help a firm retain or increase market share, and hence improve performance by increasing sales or revenues.

3.1. Productivity Model: ISO 9000 Leads to Internal Improvements

Implementing ISO 9000 forces significant discipline on the firm: it has to design procedures to ensure quality is constantly measured, procedures to ensure appropriate corrective action is taken whenever defects occur, etc. As a result, defect rates should decrease and defects should be caught earlier allowing them to be corrected at lower cost. This discipline can also identify current practices that are obsolete or counterproductive. The procedures must be sufficiently well defined that they can be documented. Once documented, they become easier to transfer to new employees, so the firm becomes less dependent on tacit knowledge residing in individual employees (Mukherjee et al. 1998). Elmuti and Kathawala (1997) report that ISO 9000 increases productivity by improving employee morale. Through all of these mechanisms, manufacturing costs should decrease when a firm starts implementing ISO 9000. We use cost of goods sold divided by sales (COGS/SALES) to measure such improvements in manufacturing productivity.

HYPOTHESIS 3. *ISO 9000 certification leads to lower COGS/SALES ratios.*

A broader view of a firm's overall productivity includes customer service processes, the cost of the certification, etc., which fall under selling, general, and administrative (SG&A) expenses. We do not report the broader cost measure (COGS + SG&A)/SALES separately because it is linked to ROS through $1 - (COGS + SG\&A)/SALES$.

3.2. Market Model: ISO 9000 Leads to Market Benefits

How does ISO 9000 certification help in the market? Certification may increase the likelihood of winning a contract, or it may even be a requirement. Alternatively, certification may help firms to reduce price pressure. In both cases, the value of sales would increase. This could occur after certification, though the improved procedures implemented prior to certification could also already positively impact sales. Hence, we use relative growth in SALES, defined by $(SALES_t - SALES_{t-1})/SALES_{t-1}$, as one of our external performance measures. To determine whether sales growth reflects firm growth or faster asset turnover, we also consider SALES/ASSETS, following Hendricks and Singhal (1997) and Easton and Jarrell (1998).

HYPOTHESIS 4. *ISO 9000 certification leads to higher relative SALES growth and to increased asset turnover (SALES/ASSETS).*

4. Data

This study is based on two large data sets. World Preferred kindly shared their database (then known as the Globus Registry) of ISO 9000 certifications in the

United States through 1997, a total of 21,482 certifications in all SIC codes at that time. This database listed the name and address of the certified site, the parent company, primary and secondary SIC code, and date of certification. World Preferred is the most widely used source of information on ISO 9000 certification activity in the United States (perhaps jointly with McGraw-Hill). Although we are not aware of inaccuracies in the World Preferred Registry, we address the consequences of possible errors below.

Firms can have multiple certifications, even within one site, while financial performance is measured at the corporate level; there exists no way of measuring financial performance of that part of a company that falls within the scope of a given ISO 9000 certification. Therefore, following Docking and Dowen (1999) and Simmons and White (1999), we focus on a firm's first certification, as the effects of certification (if any) will begin to show around that time. For small firms with a single site, the disconnect between site-level certifications and firm-level performance is not a major concern, while for large firms with many sites, this disconnect makes our tests more conservative: if the sample contains many multisite firms with only one certified site, any effects of certification must be strong in order to be detected. We return to this point in our discussion of the limitations of this study.

Our first certification occurred in 1990, so we used Compustat Annual Industrial File data for 1987–1997. Merging the two databases was a major challenge, as no unique company identifier was listed in both. We used several methods to merge them: in some easy instances, the parent company name and SIC code in the certification database matched the name and SIC code of a company listed in Compustat. In many cases, companies with similar but nonidentical names appeared in both databases; we used several automated procedures to identify such possible matches and verified them manually, using various resources such as Hoover's Online, Google, company websites, SEC disclosures, etc. In all other cases, we verified whether a site or parent company listed in the certification database was in fact a subsidiary of a differently named company in Compustat. Sites with ISO 9000 certification for which none of these procedures yielded a match in Compustat were discarded; these are likely to be foreign owned or privately held, and hence not listed in Compustat. We started with 21,482 certified sites; for 7,598 of these, we identified matching firms in Compustat, which suggests that approximately 35.4% of certified sites belong to publicly traded firms. This percentage corresponds closely to that in a recent survey of ISO 9000 and ISO 14000 certified sites in the United States (see Corbett 2004), where 67% of 933 respondents reported being foreign owned or privately held.

Altogether, merging the data took over two years. Given the scope and the manual nature of the work involved, two types of errors are theoretically possible: firms that did receive certification in 1990–1997 may be listed as noncertified in our final data set, and firms that did not receive certification may erroneously be listed as certified. In a random sample of 100 firms, we found only one firm listed as noncertified for which we were eventually able to identify a certification, and no instances of firms incorrectly labeled as certified. The latter type of error is indeed highly unlikely, as we only mark Compustat firms as certified if we have a positive match. Incorrectly labeling a Compustat firm as noncertified could occur more easily, but this would only make our tests more conservative, as long as the reliability of the matching is not correlated with the interaction between ISO 9000 certification and financial performance. If we treat some certified firms as noncertified, but still find a significant effect (as we do), we would most likely underestimate its significance. Moreover, given the very low error rate we detected, any such effects are likely to be minimal.

Another potential source of errors lies in the World Preferred Registry itself, though there are reasons to believe the database is accurate. World Preferred receives certification information directly from the registrars responsible. Registrars and World Preferred would clearly want to avoid listing certifications that have not been awarded. Conversely, one use of the World Preferred Registry is to determine the market share of registrars operating in the United States; therefore, registrars will want the World Preferred Registry to contain all certifications for which they are responsible, as otherwise their market share would be underestimated. Similarly, individual firms will want their certification to be appropriately listed. Even if some certifications are not included in the database, this would again make our tests more conservative.

We focused on manufacturing firms, i.e., all firms with SIC codes 2000–3999. In most instances, we selected control firms using two-digit SIC codes because of the higher proportion of discrepancies in three- or four-digit SIC codes (Guenther and Rosman 1994) and because the sample size otherwise shrinks rapidly. We did separate SIC code 357 (computer equipment) from the other sectors within SIC code 35 (industrial and commercial machinery, which includes engines and turbines, farm and garden machinery and equipment, etc.) as it is far more high-tech in nature and sufficiently large to treat separately.

Of the 7,238 firms listed in Compustat with SIC codes 2000–3999, 1,103 firms received at least one ISO 9000 certification during 1990–1997. We eliminated 310 observations with insufficient financial

information. We then matched each firm to a control firm or portfolio based on precertification ROA, assets, or both (as well as industry); if no suitable control firm existed for any one of the matching criteria, the observation was discarded. This was done to keep the sample constant under different matching methods, in order to avoid our comparisons being contaminated by sample effects. After that, we trimmed the data (following Barber and Lyon 1996), eliminating a further one to two outliers at each tail of the population. After trimming, we have 554 firms for which control firms exist under all matching criteria. Tables 1a and 1b give descriptive statistics for our sample.

Measuring Tobin's Q is not unambiguous; Chung and Pruitt (1994) propose approximating Tobin's Q by $(MVE + PS + DEBT)/TA$, where MVE is market value of equity, PS the liquidating value of the firm's outstanding preferred stock, DEBT the value of the

Table 1a Summary Statistics for Certified and Control Firms Using One-to-One Matching by Industry Combined with ROA and/or ASSETS (Year $t - 2$)

	N	Mean	Median	St. dev.	Min.	Max.
Certified firms						
TOTAL ASSETS[a]	554	1,629.32	218.82	5,717.09	1.57	74,293.00
SALES[a]	554	1,649.11	252.77	5,546.70	3.43	79,557.00
ROA[b]	554	15.72	15.26	8.12	−33.08	60.69
ROS[b]	554	13.28	12.52	8.30	−60.44	52.93
Tobin's Q[c]	501	1.51	1.04	1.49	0.10	14.25
COGS/SALES[b]	554	63.03	65.45	15.72	17.57	94.80
SALES/ASSETS[b]	554	127.40	121.62	45.42	18.34	395.47
Control firms (matched by industry, ROA and ASSETS)						
TOTAL ASSETS[a]	554	1,662.83	197.10	6,266.39	1.32	86,343.00
SALES[a]	554	1,749.33	266.84	7,097.30	1.70	110,496.00
ROA[b]	554	15.66	15.16	8.03	−32.14	58.16
ROS[b]	554	13.30	12.52	15.65	−295.66	48.53
Tobin's Q[c]	449	1.50	1.18	1.22	−0.83	11.58
COGS/SALES[b]	554	63.67	66.48	16.98	12.93	147.36
SALES/ASSETS[b]	554	127.08	118.65	54.20	10.87	382.25
Control firms (matched by industry, ROA)						
TOTAL ASSETS[a]	554	1,681.39	71.71	9,749.92	0.23	121,307.70
SALES[a]	554	1,577.76	96.27	8,953.79	0.27	118,571.60
ROA[b]	554	15.71	15.31	8.12	−32.14	62.82
ROS[b]	554	12.81	12.38	15.66	−295.66	42.67
Tobin's Q[c]	438	1.58	1.14	1.41	−0.07	11.58
COGS/SALES[b]	554	62.86	64.21	17.30	8.91	147.36
SALES/ASSETS[b]	554	133.08	124.34	59.88	10.87	456.15
Control firms (matched by industry, ASSETS)						
TOTAL ASSETS[a]	554	1,643.90	218.76	6,275.25	1.48	85,280.00
SALES[a]	554	1,588.92	254.20	5,755.14	1.25	78,394.00
ROA[b]	554	12.73	12.80	10.55	−32.89	58.79
ROS[b]	554	9.31	11.15	25.01	−334.40	48.10
Tobin's Q[c]	457	1.37	1.00	1.13	−0.51	9.36
COGS/SALES[b]	554	63.33	65.72	18.10	10.38	162.75
SALES/ASSETS[b]	554	119.61	114.23	53.62	2.02	372.06

[a]TOTAL ASSETS and SALES are in millions of dollars.
[b]ROA, ROS, COGS/SALES, and SALES/ASSETS are expressed in percentages.
[c]Tobin's Q is a dimensionless measure.

Table 1b Summary Statistics for Certified and Control Firms Using Portfolio Matching by Industry Combined with ROA and/or ASSETS (Year $t - 2$)

	N	Mean	Median	St. dev.	Min.	Max.
Certified firms						
TOTAL ASSETS[a]	554	1,629.32	218.82	5,717.09	1.57	74,293.00
SALES[a]	554	1,649.11	252.77	5,546.70	3.43	79,557.00
ROA[b]	554	15.72	15.26	8.12	−33.08	60.69
ROS[b]	554	13.28	12.52	8.30	−60.44	52.93
Tobin's Q[c]	501	1.51	1.04	1.49	0.10	14.25
COGS/SALES[b]	554	63.03	65.45	15.72	17.57	94.80
SALES/ASSETS[b]	554	127.40	121.62	45.42	18.34	395.47
Control firms (matched by industry, ROA and ASSETS)						
TOTAL ASSETS[a]	554	1,690.09	200.72	6,113.32	2.21	86,343.00
SALES[a]	554	1,770.13	254.16	6,941.21	2.24	110,496.00
ROA[b]	554	15.54	15.15	7.91	−31.38	58.16
ROS[b]	554	13.86	13.27	9.60	−115.14	53.68
Tobin's Q[c]	515	1.49	1.28	0.98	−0.83	5.62
COGS/SALES[b]	554	63.46	63.06	13.71	12.93	145.48
SALES/ASSETS[b]	554	125.76	121.31	44.49	20.99	382.22
Control firms (matched by industry, ROA)						
TOTAL ASSETS[a]	554	1,713.34	289.53	4,519.50	7.83	50,674.63
SALES[a]	554	1,720.15	330.71	4,750.46	5.36	53,180.35
ROA[b]	554	15.66	15.35	8.02	−32.75	60.49
ROS[b]	554	12.69	13.04	8.72	−136.12	32.72
Tobin's Q[c]	548	1.57	1.35	0.98	0.11	7.90
COGS/SALES[b]	554	62.37	61.93	10.03	32.52	91.62
SALES/ASSETS[b]	554	137.59	135.69	28.33	47.50	250.52
Control firms (matched by industry, ASSETS)						
TOTAL ASSETS[a]	554	1,679.87	212.17	5,806.56	2.16	65,985.54
SALES[a]	554	1,656.42	254.87	5,744.50	2.64	68,241.78
ROA[b]	554	10.58	11.40	7.01	−48.09	24.07
ROS[b]	554	1.33	9.73	33.81	−302.28	26.22
Tobin's Q[c]	554	1.43	1.31	0.73	0.12	4.82
COGS/SALES[b]	554	70.65	67.16	27.21	35.76	335.28
SALES/ASSETS[b]	554	117.82	121.69	25.46	55.11	242.46

[a]TOTAL ASSETS and SALES are in millions of dollars.
[b]ROA, ROS, COGS/SALES, and SALES/ASSETS are expressed in percentages.
[c]Tobin's Q is a dimensionless measure.

firm's short-term liabilities net of its short-term assets plus the book value of long-term debt, and TA the book value of total assets of the firm. Adams (1999) uses this same formula in his New Zealand study of adoption of ISO 9000, and King and Lenox (2002) use the same in their study of financial effects of various types of pollution reduction.

5. Methodology and Results

The ideal approach to test our hypotheses would be to formulate a panel data model (see, for instance, Hsiao 1993), explaining ROA in terms of certification and several control variables. A major challenge is that the certification variable is discrete and endogenous. A common approach to drawing causal inferences from nonexperimental data is Heckman's (1978) two-stage method, which first predicts the probability that a firm will seek ISO 9000 certification, based on

its characteristics, and then uses the predicted probability as independent variable in the second stage. In our case, this method has several drawbacks. First, there exists no single generally accepted extension of this approach to a panel data context, and the results we obtained were highly sensitive to the exact model specification employed. Second, with all specifications we tried, the predicted certification probabilities in the first stage were very poor predictors of actual certification, which further reduced our confidence in this approach with this particular data set. A full analysis of this panel data set would require maximum likelihood estimation with several different distributional assumptions to examine sensitivity of the results; this is left for future research.

For these reasons, we revert to event-study methods. We largely follow the approach for detecting abnormal operating performance outlined in Barber and Lyon (1996). An event study requires specifying the event period—in our case, the period during which ISO 9000 is implemented. A typical implementation takes anywhere between 6 and 18 months, so we define our event period as the year in which the certification was received (year t) and the year immediately preceding certification (year $t-1$). The period preceding the event period (year $t-2$) is used for determining the control group. The three years following the event period (years $t+1$ through $t+3$) are used to test for abnormal performance.

Barber and Lyon (1996) discuss three key choices in designing an event study. First, the selection of a measure of operating performance: they find that operating income divided by assets, or ROA is the preferred measure (compared to other similar measures of profitability). We report results on ROA and ROS. The second decision is to specify a firm's predicted performance in the absence of the event. This is generally done by comparing the certified firms' postevent performance with a control group of one or more firms, normally drawn from the same industry (two-digit SIC code), that did not experience the event. Barber and Lyon (1996) show that selecting a control group with similar pre-event performance to the certified firm drastically increases the power of the study, especially when the certified firms already have higher pre-event performance, as is the case with ISO 9000. Lie (2001) confirms that matching by ROA is appropriate in this case.

Although Barber and Lyon (1996) find that matching by size in addition to industry and pre-event performance provides little or no extra benefit, we report this matching as our main results, given that our Compustat sample does include small firms listed on NASDAQ as well as larger firms listed on the NYSE and other exchanges. The control firm is the

firm in the same industry that is "closest" to the certified firm in terms of assets and ROA. We transformed assets and ROA in each year into z-scores by subtracting the industry mean and dividing by the industry standard deviation for that year; "closest" was then defined by the smallest Euclidean distance in the two-dimensional space of z-scores, subject to the constraint that the control firm's ROA had to lie between 90% and 110% of the certified firm's ROA, and its assets had to lie between 50% and 200% of the certified firm's assets. Barber and Lyon (1996) used the same range for ROA, but only a 70%–130% range for assets; we expanded the range in response to a reviewer's suggestion, as this reduces the number of certified firms for which no match can be found. Expanding the matching range further, for instance to a factor three from the certified firm's assets, as Hendricks and Singhal (1997) do, yielded very few additional observations in our case. Many other studies, including Hendricks and Singhal (1997), select control firms based on industry and size alone; we report those results too, as well as results based on matching by industry and performance alone. The control group can consist of a single closely matching firm, as in Hendricks and Singhal (1997), or of a portfolio of closely matching firms, as in Barber and Lyon (1996); we report results using both methods. Control portfolios consist of all firms that meet the above matching criteria, rather than just the closest match.

There is no unique best way to choose a control sample. Another matching method uses propensity scores, but Smith and Todd (2005) find that the evidence to date about its superior performance is questionable. The control firms should be as similar as possible to the firms receiving ISO 9000 certification, so matching on multiple dimensions is preferred; this is why the matching by industry, ROA, and ASSETS provides our main results.

Expected performance $E[P_{i,\tau,l}]$ of firm i in any period $\tau + l$, using period τ as base period, is given by $E[P_{i,\tau,l}] = P_{i,\tau} + (PI_{i,\tau+l} - PI_{i,\tau})$, where $PI_{i,\tau}$ is the performance of firm i's control group in period τ. Abnormal performance is then given by $AP_{i,\tau,l} = P_{i,\tau+l} - E[P_{i,\tau,l}]$, and the event study consists of testing whether this abnormal performance is significantly different from zero in the hypothesized direction.

Barber and Lyon's (1996) third finding is that, in their study, nonparametric tests are more powerful than parametric (t-) tests. The choice of a test statistic depends on the distribution of abnormal performance; in our case, the Shapiro-Wilk test always rejected the hypothesis that abnormal performance was normally distributed with p-values below 0.01, indicating that nonparametric tests are called for. When the distribution is symmetric we can use the

Wilcoxon signed-rank test; in other cases, the distribution is highly skewed so we use the less powerful sign test (Conover 1999). For completeness, we report all three statistics.

We can now test the hypotheses by examining whether certified firms experience abnormal performance in terms of ROA, ROS, Tobin's Q, COGS/SALES, SALES, and SALES/ASSETS, respectively. Tables 2–7 show the results for ROA using all six matching criteria. Tables 8–12 show the results for the remaining performance measures, using one-to-one matching by industry, ROA, and size. Table 13 summarizes the results for ROA across all six matching methods. The full results are available in a separate document. For each test, we report the sample size under "N." The next two columns report the

Table 2 Abnormal Performance in ROA; One-to-One Matching, by Industry, ROA, ASSETS

From year	N[a]	AP mean[b]	AP median[b]	sk[c]	p-value (t-test)[d]	p-value (WSR test)[d]	p-value (sign test)[d]
$t-3$ to $t-2$	405	0.742	0.423		0.0695*	0.0595*	0.0745*
$t-2$ to $t-1$	554	0.620	0.124		0.0824*	0.0766*	0.2220
$t-1$ to t	554	1.508	0.584	s	0.0075**	0.0017***	0.0167**
t to $t+1$	363	−0.288	−0.246		0.6913	0.7809	0.7847
$t+1$ to $t+2$	233	0.417	0.469		0.2219	0.1151	0.0839*
$t+2$ to $t+3$	147	−0.525	0.108		0.7594	0.6171	0.4345
$t-2$ to t	554	2.128	0.996		0.0009***	0.0002***	0.0097***
$t-2$ to $t+1$	363	1.911	1.035		0.0014**	0.0006***	0.0090***
$t-2$ to $t+2$	233	3.183	2.437		0.0001***	0.0000***	0.0000***
$t-2$ to $t+3$	147	3.936	2.575	s	0.0001***	0.0001***	0.0008***
$t-1$ to $t+1$	363	1.343	0.953		0.0177**	0.0296**	0.0090***
$t-1$ to $t+2$	233	2.230	2.597		0.0040***	0.0001***	0.0000***
$t-1$ to $t+3$	147	3.575	3.371	s	0.0010***	0.0002***	0.0158**

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$. In all cases, the hypothesis of normally distributed abnormal performance was rejected by the Shapiro-Wilk test with p-values below 1%, so the t-test is never appropriate, but reported for completeness; if the skewness is substantial the sign test is appropriate, otherwise one can use the WSR test.

[a]"N" indicates total number of certified firms in the corresponding sample.

[b]AP mean and median indicate the mean and median abnormal performance.

[c]An "s" in column "sk" indicates that the absolute value of the skewness is greater than one.

[d]The p-values shown are those for the one-sided test of the null hypothesis of no abnormal performance, using a t-test, Wilcoxon signed rank test, and sign test respectively.

Table 3 Abnormal Performance in ROA; Portfolio Matching, by Industry, ROA, ASSETS

From year	N	AP mean	AP median	sk	p-value (t-test)	p-value (WSR test)	p-value (sign test)
$t-3$ to $t-2$	405	−0.0173	0.0000		0.5161	0.4905	0.5000
$t-2$ to $t-1$	554	0.8906	0.9124		0.0102**	0.0001***	0.0002***
$t-1$ to t	554	1.6709	1.3656	s	0.0018***	0.0000***	0.0000***
t to $t+1$	363	0.2498	0.3803		0.3197	0.0990*	0.0944*
$t+1$ to $t+2$	233	0.5664	0.3695		0.0813*	0.0738*	0.1624
$t+2$ to $t+3$	147	0.0851	0.6468		0.4519	0.3445	0.1241
$t-2$ to t	554	2.5615	1.6100		0.0000***	0.0000***	0.0000***
$t-2$ to $t+1$	363	2.8874	1.8706		0.0000***	0.0000***	0.0000***
$t-2$ to $t+2$	233	3.7216	1.9982		0.0000***	0.0000***	0.0000***
$t-2$ to $t+3$	147	4.1047	2.9681	s	0.0000***	0.0000***	0.0000***
$t-1$ to $t+1$	363	2.3241	1.6970	s	0.0000***	0.0000***	0.0000**
$t-1$ to $t+2$	233	2.9311	2.1526		0.0000***	0.0000***	0.0000***
$t-1$ to $t+3$	147	3.5600	2.9978	s	0.0003***	0.0001***	0.0103**

Table 4 Abnormal Performance in ROA; One-to-One Matching, by Industry, ROA

From year	N	AP mean	AP median	sk	p-value (t-test)	p-value (WSR test)	p-value (sign test)
$t-3$ to $t-2$	405	−0.926	−0.772		0.9359	0.9698	0.9104
$t-2$ to $t-1$	554	2.666	1.301	s	0.0000***	0.0000***	0.0000***
$t-1$ to t	554	2.197	0.905		0.0005***	0.0001***	0.0042***
t to $t+1$	363	0.978	0.028	s	0.0884*	0.2114	0.3964
$t+1$ to $t+2$	233	−0.585	0.171	s	0.7070	0.8539	0.3230
$t+2$ to $t+3$	147	1.701	0.444		0.0750*	0.0619*	0.0346**
$t-2$ to t	554	4.863	2.620		0.0000***	0.0000***	0.0000***
$t-2$ to $t+1$	363	5.633	2.824	s	0.0000***	0.0000***	0.0000***
$t-2$ to $t+2$	233	5.714	2.688	s	0.0000***	0.0000***	0.0000***
$t-2$ to $t+3$	147	8.976	3.751	s	0.0000***	0.0000***	0.0040***
$t-1$ to $t+1$	363	3.424	1.364		0.0004***	0.0002***	0.0118**
$t-1$ to $t+2$	233	3.166	2.172		0.0123**	0.0013***	0.0035***
$t-1$ to $t+3$	147	6.965	4.604		0.0006***	0.0001***	0.0065***

Table 5 Abnormal Performance in ROA; Portfolio Matching, by Industry, ROA

From year	N	AP mean	AP median	sk	p-value (t-test)	p-value (WSR test)	p-value (sign test)
t − 3 to t − 2	405	−1.338	−1.304		0.9998	1.0000	1.0000
t − 2 to t − 1	554	2.314	2.377		0.0000***	0.0000***	0.0000***
t − 1 to t	554	1.395	1.710	s	0.0004***	0.0000***	0.0000***
t to t + 1	363	0.552	0.573	s	0.1303	0.0219**	0.0293**
t + 1 to t + 2	233	0.936	0.945		0.0108**	0.0035***	0.0179**
t + 2 to t + 3	147	0.888	1.179		0.0328**	0.0048***	0.0008***
t − 2 to t	554	3.709	3.550	s	0.0000***	0.0000***	0.0000***
t − 2 to t + 1	363	4.577	3.934		0.0000***	0.0000***	0.0000***
t − 2 to t + 2	233	5.760	4.458		0.0000***	0.0000***	0.0000***
t − 2 to t + 3	147	7.039	5.037		0.0000***	0.0000***	0.0000***
t − 1 to t + 1	363	2.438	2.543		0.0000***	0.0000***	0.0000***
t − 1 to t + 2	233	3.632	3.504		0.0000***	0.0000***	0.0000***
t − 1 to t + 3	147	5.280	4.353		0.0000***	0.0000***	0.0000***

Table 7 Abnormal Performance in ROA; Portfolio Matching, by Industry, ASSETS

From year	N	AP mean	AP median	sk	p-value (t-test)	p-value (WSR test)	p-value (sign test)
t − 3 to t − 2	405	0.8539	0.9685		0.0103**	0.0021***	0.0019***
t − 2 to t − 1	554	0.5841	0.6719		0.0385**	0.0006***	0.0029***
t − 1 to t	554	0.2966	0.4386		0.2204	0.0148**	0.0804*
t to t + 1	363	0.1412	0.1225	s	0.3735	0.3820	0.3373
t + 1 to t + 2	233	0.3289	0.6346		0.1481	0.0295**	0.0578*
t + 2 to t + 3	147	0.6515	1.0701		0.0727*	0.0204**	0.0103**
t − 2 to t	554	0.8807	1.1045	s	0.0251**	0.0001***	0.0009***
t − 2 to t + 1	363	0.7244	0.3369		0.0605*	0.0353**	0.1241
t − 2 to t + 2	233	1.3506	1.1356		0.0047***	0.0035***	0.0179**
t − 2 to t + 3	147	2.3401	1.1860		0.0005***	0.0018***	0.0494**
t − 1 to t + 1	363	0.5884	0.4219		0.0812*	0.0906*	0.1469
t − 1 to t + 2	233	1.0446	0.9561		0.0129**	0.0039***	0.0019***
t − 1 to t + 3	147	2.1729	2.2921		0.0007***	0.0007***	0.0040***

Table 6 Abnormal Performance in ROA; One-to-One Matching, by Industry, ASSETS

From year	N	AP mean	AP median	sk	p-value (t-test)	p-value (WSR test)	p-value (sign test)
t − 3 to t − 2	405	0.530	0.559		0.1542	0.1026	0.0896*
t − 2 to t − 1	554	0.232	0.073		0.2997	0.4058	0.3353
t − 1 to t	554	0.561	0.097		0.1710	0.4231	0.4325
t to t + 1	363	1.029	0.315	s	0.0866*	0.2375	0.3181
t + 1 to t + 2	233	−0.709	0.384		0.8817	0.4644	0.2351
t + 2 to t + 3	147	2.766	0.800	s	0.0025***	0.0112**	0.0103**
t − 2 to t	554	0.794	−0.151		0.1307	0.3626	0.7380
t − 2 to t + 1	363	2.026	−0.533	s	0.0204**	0.2535	0.7847
t − 2 to t + 2	233	1.847	2.300		0.0255**	0.0208**	0.0654*
t − 2 to t + 3	147	5.825	3.610	s	0.0001***	0.0002***	0.0005***
t − 1 to t + 1	363	1.850	0.150	s	0.0159**	0.2526	0.3565
t − 1 to t + 2	233	1.384	1.097		0.0578*	0.0407**	0.0283**
t − 1 to t + 3	147	5.021	2.264	s	0.0005***	0.0026***	0.0158**

Table 8 Abnormal Performance in ROS; One-to-One Matching, by Industry, ROA, ASSETS

From year	N	AP mean	AP median	sk	p-value (t-test)	p-value (WSR test)	p-value (sign test)
t − 3 to t − 2	405	0.266	0.024		0.2340	0.3584	0.4603
t − 2 to t − 1	554	0.442	0.347	s	0.2113	0.0097***	0.0306**
t − 1 to t	554	1.074	0.255		0.0053***	0.0030***	0.0252**
t to t + 1	363	0.201	−0.058		0.3199	0.4565	0.6036
t + 1 to t + 2	233	0.887	0.604	s	0.0579**	0.0910*	0.0208**
t + 2 to t + 3	147	−0.177	0.245		0.6121	0.3175	0.3396
t − 2 to t	554	1.516	0.973	s	0.0034***	0.0001***	0.0022***
t − 2 to t + 1	363	2.024	0.810		0.0003***	0.0001***	0.0155**
t − 2 to t + 2	233	3.343	1.792		0.0000***	0.0000***	0.0004***
t − 2 to t + 3	147	4.121	2.788		0.0000***	0.0000***	0.0000***
t − 1 to t + 1	363	1.161	0.498		0.0114**	0.0081***	0.0201**
t − 1 to t + 2	233	2.161	1.682		0.0026***	0.0002***	0.0004***
t − 1 to t + 3	147	3.275	2.069		0.0002***	0.0000***	0.0000***

Table 9 Abnormal Performance in Tobin's Q; One-to-One Matching, by Industry, ROA, ASSETS

From year	N	AP mean	AP median	sk	p-value (t-test)	p-value (WSR test)	p-value (sign test)
$t-3$ to $t-2$	282	0.031	−0.002		0.3321	0.7853	0.5711
$t-2$ to $t-1$	406	0.083	0.052		0.0973*	0.0211**	0.0368**
$t-1$ to t	468	0.097	0.065		0.0315**	0.0089***	0.0080***
t to $t+1$	315	−0.020	0.018		0.6295	0.4997	0.1987
$t+1$ to $t+2$	200	−0.076	−0.003		0.8763	0.4930	0.7146
$t+2$ to $t+3$	126	−0.122	−0.067		0.9792	0.9606	0.9634
$t-2$ to t	405	0.186	0.179		0.0069***	0.0002***	0.0000***
$t-2$ to $t+1$	256	0.167	0.209		0.0248**	0.0011***	0.0003***
$t-2$ to $t+2$	160	0.212	0.200		0.0375**	0.0015***	0.0034***
$t-2$ to $t+3$	102	0.195	0.179		0.0436**	0.0308**	0.0364**
$t-1$ to $t+1$	298	0.057	0.057		0.2176	0.1279	0.0656*
$t-1$ to $t+2$	182	0.061	0.090	s	0.2870	0.0208**	0.0265**
$t-1$ to $t+3$	116	0.149	0.164		0.0738*	0.0221**	0.0124**

Table 10 Abnormal Performance in COGS/SALES; One-to-One Matching, by Industry, ROA, ASSETS

From year	N	AP mean	AP median	sk	p-value (t-test)	p-value (WSR test)	p-value (sign test)
$t-3$ to $t-2$	405	−0.432	−0.217	s	0.1155	0.1140	0.0970*
$t-2$ to $t-1$	554	−0.215	−0.398	s	0.2566	0.0191**	0.0067***
$t-1$ to t	554	−0.812	−0.242	s	0.0055***	0.0253**	0.0625*
t to $t+1$	363	−0.240	−0.299		0.2011	0.1052	0.0413**
$t+1$ to $t+2$	233	−0.848	−0.404	s	0.0367**	0.1380	0.0654*
$t+2$ to $t+3$	147	0.349	0.444		0.7561	0.7076	0.8928
$t-2$ to t	554	−1.027	−0.533		0.0037***	0.0088***	0.0107**
$t-2$ to $t+1$	363	−1.328	−1.143		0.0024***	0.0006***	0.0067***
$t-2$ to $t+2$	233	−2.503	−1.559		0.0002***	0.0000***	0.0000***
$t-2$ to $t+3$	147	−3.283	−2.129		0.0000***	0.0000***	0.0001***
$t-1$ to $t+1$	363	−0.890	−0.592		0.0135**	0.0077***	0.0779*
$t-1$ to $t+2$	233	−1.803	−1.163		0.0026***	0.0010***	0.0006***
$t-1$ to $t+3$	147	−2.348	−1.460		0.0014***	0.0006***	0.0065***

Table 11 Abnormal Relative SALES Growth; One-to-One Matching, by Industry, ROA, ASSETS

From year	N	AP mean	AP median	sk	p-value (t-test)	p-value (WSR test)	p-value (sign test)
$t-3$ to $t-2$	405	−0.023	−0.011		0.8829	0.9858	0.8847
$t-2$ to $t-1$	554	0.010	−0.018		0.2976	0.8068	0.9427
$t-1$ to t	554	0.037	0.014		0.0049***	0.0119**	0.0625*
t to $t+1$	363	0.049	0.015		0.0071***	0.0082***	0.0460**
$t+1$ to $t+2$	233	0.006	0.019		0.4056	0.1911	0.1785
$t+2$ to $t+3$	147	0.044	0.034		0.0225**	0.0347**	0.0673*
$t-2$ to t	554	0.070	0.000		0.0303**	0.2785	0.5000
$t-2$ to $t+1$	363	0.091	0.031		0.0965*	0.0301**	0.1342
$t-2$ to $t+2$	233	0.118	0.016		0.1038	0.1205	0.3466
$t-2$ to $t+3$	147	0.220	0.049	s	0.1099	0.1585	0.2533
$t-1$ to $t+1$	363	0.090	0.028		0.0099***	0.0018***	0.0938*
$t-1$ to $t+2$	233	0.116	0.049		0.0291**	0.0280**	0.1462
$t-1$ to $t+3$	147	0.236	0.116	s	0.0192**	0.0253**	0.0919*

Notes. "Abnormal relative sales growth" is defined by comparing $(SALES_t - SALES_{t-1})/SALES_{t-1}$ of certified firms with that of the control firms.

mean and median abnormal performance. The first row reports abnormal performance (AP) in the year prior to the matching (up to two years before certification), i.e., $AP_{i,\tau,l}$ with $\tau = t-3$ and $l = 1$. (We discuss later why this was included.) The second row reports abnormal performance in the year immediately following the decision to embark on the certification process—i.e., $AP_{i,\tau,l}$ with $\tau = t-2$ and $l = 1$—the year before the certification is awarded. The third row measures abnormal performance in the year of certification, i.e., $AP_{i,\tau,l}$ with $\tau = t-1$ and $l = 1$, etc. The second and third block of results report longer-term abnormal performance; for instance, the row "−2 to 0" measures $AP_{i,\tau,l}$ with $\tau = t-2$ and $l = 2$. The column marked "sk" shows an "s" if the absolute skewness of the distribution of abnormal performance, defined as the third moment of the distribution divided by the cube of the standard deviation, is greater than one; Hair et al. (1998) suggest that indicates severe skewness. The next three columns report the *p*-values corresponding to parametric and nonparametric tests of the hypothesis that certified firms exhibit superior performance (i.e., positive abnormal performance for ROA, ROS, Tobin's Q, SALES, and SALES/ASSETS and negative abnormal performance for COGS/SALES).

Table 12 Abnormal Performance in SALES/ASSETS; One-to-One Matching, by Industry, ROA, ASSETS

From year	N	AP mean	AP median	sk	p-value (t-test)	p-value (WSR test)	p-value (sign test)
t − 3 to t − 2	405	2.468	0.891		0.0765*	0.1579	0.1850
t − 2 to t − 1	554	−2.511	−0.947		0.9648	0.8630	0.8463
t − 1 to t	554	3.486	3.472		0.0039***	0.0004***	0.0008***
t to t + 1	363	−1.506	−0.863		0.8252	0.8400	0.9221
t + 1 to t + 2	233	−3.674	−0.555		0.9672	0.7120	0.6770
t + 2 to t + 3	147	−4.926	−2.601		0.9956	0.9943	0.9717
t − 2 to t	554	0.975	1.097		0.2636	0.1619	0.1344
t − 2 to t + 1	363	−0.500	0.250		0.5935	0.4675	0.3369
t − 2 to t + 2	233	−2.465	0.004		0.7836	0.4360	0.4738
t − 2 to t + 3	147	−6.796	−3.309	s	0.9436	0.8368	0.7187
t − 1 to t + 1	363	2.152	2.156		0.1256	0.1042	0.0413**
t − 1 to t + 2	233	−1.198	2.428		0.6666	0.2602	0.3230
t − 1 to t + 3	147	−1.833	−2.469		0.6967	0.4527	0.7187

6. Discussion and Limitations

6.1. Discussion of Results

In this section, we discuss our main results, after which we add some notes about interpretation. Table 2 shows that when selecting single control firms based on industry, ROA, and size, firms experienced significant abnormal improvements in ROA in the year of their first ISO 9000 certification. Using the sign test, the most conservative of the three, abnormal performance for $t − 1$ to t, the year leading up to certification, is weakly significant ($p = 0.0167$). After certification, no further abnormal improvements in ROA occur. The cumulative improvements, however, appear to be strong and lasting, as all the longer-term tests, using either $t − 2$ or $t − 1$ as base period, show significant abnormal improvements in ROA, usually with p-values below 0.01. The magnitude of the longer-term effects is considerable, in the order of 1–4 percentage points. Starting with an ROA of 15.7% (the mean of certified firms in year $t − 2$, from Table 1a), the mean abnormal increase in ROA by year $t + 3$ is 3.94 percentage points (Table 2), compared to noncertified firms starting with the same ROA, which corresponds to a relative increase of 25%.

Tables 3–5 show that matching by industry and ROA alone, and/or using portfolios of control firms, has little effect on the long-term results, although the pure performance-matched sample exhibits a level of abnormal performance that is almost twice as high. The short-term effects do become much stronger in all three cases, in the year of certification ($t − 1$ to t) and now also in the preceding year ($t − 2$ to $t − 1$). Matching by size alone using portfolios of control firms also leads to similar results, as shown in Table 7. However, one-to-one matching based on size alone (Table 6) makes the short-term effects disappear; even there, though, the long-term effects remain strong. We discuss the difference between the methods later.

Table 13 Significance of Abnormal Performance in ROA Under All Six Matching Criteria

		Industry, ROA, ASSETS				Industry, ROA				Industry, ASSETS			
		One-to-one		Portfolio		One-to-one		Portfolio		One-to-one		Portfolio	
From year	N[a]	WSR[b]	sign[b]	WSR[b]	sign[b]	WSR[b]	sign[b]	WSR[b]	sign[b]	WSR[b]	sign[b]	WSR[b]	sign[b]
t − 3 to t − 2	405	*	*								*	***	***
t − 2 to t − 1	554	*		***	***	***	***	***	***			***	***
t − 1 to t	554	***	**	***	***	***	***	***	***			**	*
t to t + 1	363			*	*			**	**				
t + 1 to t + 2	233		*	*				***	**			**	*
t + 2 to t + 3	147					*	**	***	***	**	**	**	**
t − 2 to t	554	***	***	***	***	***	***	***	***			***	***
t − 2 to t + 1	363	***	***	***	***	***	***	***	***	**			
t − 2 to t + 2	233	***	***	***	***	***	***	***	***	**	*	***	**
t − 2 to t + 3	147	***	***	***	***	***	***	***	***	***	***	***	**
t − 1 to t + 1	363	**	***	***	***	***	**	***	***			*	
t − 1 to t + 2	233	***	***	***	***	***	***	***	***	**	**	***	***
t − 1 to t + 3	147	***	**	***	**	***	***	***	***	***	**	***	***

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

[a]"N" indicates total number of certified firms in the corresponding sample.

[b]The significance levels shown are those for the one-sided test of the null hypothesis of no abnormal performance, using the Wilcoxon signed rank test and sign test, respectively.

Table 13 summarizes the results for ROA across all six matching methods.

These results are very closely mirrored using ROS as performance measure, as shown in Table 8 for the one-to-one performance- and size-matched sample. Moving to Tobin's Q, in Table 9, we find a largely similar pattern, with the notable difference that the long-term effects are not as persistent across matching methods as is the case for ROA (see Tables A14–A19 in the online appendix, available at http://mansci. pubs.informs.org/ecompanion.html).

What drives the improvement in ROA? Immediately after deciding to seek certification, firms experience a significant abnormal productivity improvement, as witnessed by the drop in COGS/SALES in Table 10, which is maintained over all the longer-horizon tests. Certified firms on average see COGS/SALES decrease by 3.28 percentage points more than noncertified firms by year $t+3$ (Table 10), a relative improvement of 5% from the mean COGS/SALES ratio in $t-2$ of 63% (Table 1a). Table 11 shows a temporary weakly significant abnormal boost in SALES growth, mainly in the years immediately before and after certification. The mean (and median) magnitude of abnormal sales growth continues to increase for longer horizons but becomes less significant. Table 12 shows a significant abnormal increase in SALES/ASSETS in the year of certification, but no long-term effects because the certified firms' assets also grew faster than those of noncertified firms. The magnitude of long-term abnormal performance in SALES/ASSETS can be both positive or negative depending on the matching method, but it is significant (and positive) only for the pure performance-matched sample using control portfolios (see Tables A32–A37 in the online appendix). Certification is associated with faster growth, but does not increase firms' sales relative to their asset base. To summarize, firms' first ISO 9000 certification did lead to improvements in ROA, primarily through increased productivity.

These findings are largely consistent with those in Hendricks and Singhal (1997) for TQM. How should one interpret our results? Strictly speaking, they show that firms experience significantly better relative performance, after deciding to seek their first ISO 9000 certification, than control firms with similar assets and performance prior to that decision. This strongly suggests that after deciding to seek their first ISO 9000 certification, firms made changes that led to improvements in ROA through increased productivity. Closer examination of certified firms' performance over time suggests that in many cases, certified firms see only a minor increase in ROA compared to year $t-2$ levels, while noncertified firms see substantial declines. This would mean that ISO 9000 is a qualifying criterion, in

Hill's (1993) terms, rather than an order-winning criterion. It helps firms to maintain, rather than increase, profitability. This is no surprise: it is hard to see how firms could derive sustainable competitive advantage from a public standard like ISO 9000. However, as incomplete availability of financial data causes the sample to shrink for longer horizons, we need to exercise caution in interpreting this observation.

Are these improvements a direct causal effect of deciding to seek ISO 9000? Certainly, it is possible that the decision to seek ISO 9000 is associated with other "good management" practices, and that it is these latter practices that improve ROA rather than the ISO 9000 certification process itself. However, the control groups do consist of firms with the *same* prior performance, so something changed at the certified firms in the two years before the actual certification. In light of the magnitude of the effects, it seems likely that factors other than ISO 9000 certification contributed, but given the design of the study, these findings do indicate that the preparations for the first ISO 9000 certification also contributed to superior performance. We explore various alternative explanations below.

6.2. Effect of Matching Method and Discussion of Alternative Explanations

The results for abnormal performance in ROA using all six different matching criteria are summarized in Table 13; Tables A38–A43 in the online appendix do the same for all performance measures. When we match by ROA only, the magnitudes and significance of abnormal performance are the largest. Matching by size in addition to ROA leads to a decrease in magnitude and significance, but no noteworthy change. However, matching only by size, the short-term effects disappear under one-to-one matching (but not under portfolio matching) and the long-term effects become weaker. This is consistent with Barber and Lyon's (1996) finding that performance matching increases the power of the study and that matching by size is unimportant compared to performance-matching (p. 385), and with Hendricks and Singhal's (1997, p. 1265) report that matching by size has little impact on industry-matched results.

Although these results are quite strong in themselves, two alternative explanations still need to be addressed. The first is the possibility that the sample firms were already increasing their ROA relative to the control firms, even before deciding to seek ISO 9000 certification. If this were true, one would find positive abnormal performance in the period $t-3$ to $t-2$. One of the six matching methods (size-matched control portfolios) shows significant positive abnormal performance for $t-3$ to $t-2$. Two others (size- and ROA-matched control firms, and size-matched control firms) show weakly significant abnormal

performance ($p < 0.1$); in the first of these, the postcertification abnormal performance is much larger and more significant. The two ROA-matched samples (one-to-one and portfolios) both show strongly negative abnormal performance prior to matching. From this we conclude that the non-performance-matched samples display only a minimal tendency towards positive abnormal performance prior to the certification decision, and that this effect is too small and inconsistent to fully explain the longer-term abnormal performance observed after ISO 9000 certification.

Another potential alternative explanation for our results could be that the performance matching induces an artificially positive abnormal performance by reversion to the mean in operating performance measures. Firms that decide to seek ISO 9000 certification already have a higher ROA prior to the certification decision (as shown in Tables 1a and 1b), so requiring control firms to have equally high prior performance may lead to selecting noncertifying firms with uncharacteristically high performance in the year of the matching, but lower performance before and after. If this were true, the control firms should display positive abnormal performance over the period $t - 3$ to $t - 2$, prior to the matching, meaning that the sample firms should display an equivalent negative abnormal performance over that same period. The two ROA-matched samples do display this effect; however, the sample firms' negative abnormal performance over $t - 3$ to $t - 2$ is considerably smaller than the positive abnormal performance in each of the following two periods. Matching by size as well as ROA eliminates this effect. This suggests that there is some degree of reversion to the mean in pure ROA-matched samples, but not in the others, so that this effect is far from sufficient to explain the positive abnormal performance associated with ISO 9000 certification.

An event study can never rule out other sources of unobserved heterogeneity, and indeed, we know that firms that will decide to seek ISO 9000 certification already perform better than those that will not. The key question is, is there any unobserved heterogeneity that affects the impact of ISO 9000 on a firm's performance? The two tests above indicate that whatever unobserved heterogeneity is present is insufficient to fully explain the positive link between abnormal performance and ISO 9000 certification that we find here. One way to look at these results is to conclude that the prescriptions contained in ISO 9000 themselves lead to superior performance. Alternatively, firms may use the certification procedure as an opportunity for process improvement. In the latter view, it is the firm's own efforts that lead to superior performance, not the standard in itself, but these efforts are triggered and guided by the ISO 9000 certification process. One firm

may have the potential to be more successful than another, but this potential is turned into reality as a result of ISO 9000. One could say that some firms may have a "gene for good management," but that this gene is not expressed until ISO 9000 switches it on. (We are grateful to Paul Kleindorfer for this interpretation.) This latter view could go a long way to reconciling the apparently conflicting views of those who believe that ISO 9000 is too minimal a standard to be able to drive real performance improvements, and those who are convinced that it does just that. If a firm lacks the "good management gene," ISO 9000 may not help them much either. Firms that do have the potential to be successful may need an external trigger to unlock that potential; ISO 9000 appears to be one such trigger.

6.3. Limitations and Future Research

Do our results depend on the degree to which firms have already adopted TQM? Several interactions are possible: Firms with a history of TQM may have less room left for improvements and, hence, reap less benefits from ISO 9000 certification. Conversely, firms with a TQM program may be able to obtain ISO 9000 certification at lower cost and, hence, reap higher benefits. Evidence on the interaction between ISO 9000 and TQM is scarce, and our study does not address TQM. The survey by Terziovski et al. (1997) found no link between ISO 9000 certification and improved financial performance in Australia and New Zealand, irrespective of the degree of TQM implementation. To our knowledge, there is no hard evidence that TQM facilitates or hinders ISO 9000 certification, or vice versa, but this needs to be studied further.

What do our results imply for firms outside the United States? One reviewer suggested that firms in Europe were quicker to adopt ISO 9000 certification, while firms in the United States may have been quicker to adopt TQM practices; would the effects of ISO 9000 in Europe be similar to what we found for the United States? Again, more evidence on interactions between TQM and ISO 9000 would be needed to determine whether these effects depend on a country's prior history with TQM.

There are several other inevitable limitations. First, given the effort involved in matching the database of ISO 9000 certifications to Compustat, it was not possible to include certifications after 1997. In fact, the ISO 9000 series was thoroughly revised in 2000, which is likely to introduce a discontinuity in the effect of certification around 1998–1999, as firms rush to seek certification before the new, tougher standard is introduced. It would be interesting to replicate this study with an even longer time horizon in the future. Second, the study matches ISO 9000 certifications, which involve individual sites, with corporate

financial performance. This is not a major problem as it makes our tests more conservative, but if reliable disaggregated performance data on individual sites become available, it would be worthwhile to replicate this study for individual sites. Third, the study only includes U.S. manufacturing firms, so the findings need not apply to nonmanufacturing firms or to non-U.S. firms. Fourth, an event study can never rule out all alternative explanations. For instance, firms with higher exports to Europe are more likely to seek ISO 9000 certification (Anderson et al. 1999). If purchases by European firms grew faster than those by firms in the United States and elsewhere, during the full horizon of our study, U.S. firms exporting to Europe would presumably exhibit superior performance, which would then be correlated to ISO 9000 certification. However, we are not aware of any reason to believe that purchases from the United States by firms in Europe outpaced those by firms elsewhere during the 1990–1997 period. Moreover, in many of our tests the initial occurrence of superior performance coincided in time with the years immediately leading up to each individual firm's certification. Such arguments make it highly unlikely that such an export-focused or any other analogous explanation is the full story either. Finally, this study has only included ISO 9000 certifications; it would be interesting to replicate this study focusing on QS 9000, ISO 14000, SA 8000, and other comparable current or proposed process standards.

7. Conclusions

Using event-study methods, we have found that after deciding to seek their first ISO 9000 certification, publicly traded manufacturing firms in the United States experience significant abnormal improvements in financial performance. We find that the precise timing and magnitude of the effects vary across different control-group specifications, but the longer-term effects are strongly significant in all cases. The degree to which the control group specification affects the short-term results indicates the importance of including extensive sensitivity analyses in any event study, e.g., using several combinations of industry, size, and pre-event performance as matching criteria, and using both individual firms and portfolios as control groups. By using control firms with similar pre-event size and performance, and by performing extensive sensitivity analyses, we can largely rule out the alternative hypothesis that ISO 9000 has no causal effect on performance but is simply adopted by firms that are better managed. Of course, this does not mean that all firms will reap benefits from ISO 9000 certification: anecdotal evidence abounds that some firms implement the standard in a more rigorous and comprehensive way than others, and hence are likely to

obtain more benefits. Overall, though, the evidence presented here does support the view that careful design and implementation of consistent and documented quality management systems can contribute significantly to superior financial performance.

An online appendix to this paper is available at http://mansci.pubs.informs.org/ecompanion.html.

References

Adams, M. 1999. Determinants of ISO accreditation in the New Zealand manufacturing sector. *Omega* **27** 285–292.

Anderson, S. W., J. D. Daly, M. F. Johnson. 1999. Why firms seek ISO 9000 certification: Regulatory compliance or competitive advantage? *Production Oper. Management* **8**(1) 28–43.

Andrews, R. N. L., J. Charm, H. Habicht, T. Knowlton, M. Sale, V. Tschinkel. 2001. *Third-Party Auditing of Environmental Management Systems: U.S. Registration Practices for ISO 14001.* National Academy of Public Administration.

Barber, B. M., J. D. Lyon. 1996. Detecting abnormal operating performance: The empirical power and specification of test statistics. *J. Financial Econom.* **41** 359–399.

Casadesús, M., G. Giménez, I. Heras. 2001. Benefits of ISO 9000 implementation in Spanish industry. *Eur. Bus. Rev.* **3**(6) 327–335.

Chung, K. H., S. W. Pruitt. 1994. A simple approximation of Tobin's Q. *Financial Management* **23**(3) 70–74.

Conover, W. J. 1999. *Practical Nonparametric Statistics*, 3rd ed. John Wiley & Sons, New York.

Corbett, C. J. 2004. Diffusion of ISO 9000 and ISO 14000 through global supply chains. Working paper, Anderson School of Management, UCLA, Los Angeles, CA.

Corbett, C. J., D. A. Kirsch. 2001. International diffusion of ISO 14000 certification. *Production Oper. Management* **10**(3) 327–342.

Docking, D. S., R. J. Dowen. 1999. Market interpretation of ISO 9000 registration. *J. Financial Res.* **XXII**(2) 147–160.

Easton, G. S., S. L. Jarrell. 1998. The effects of total quality management on corporate performance: An empirical investigation. *J. Bus.* **71**(2) 253–307.

Elmuti, D., Y. Kathawala. 1997. An investigation into the effects of ISO 9000 on participants' attitudes and job performance. *Production Inventory Management J.* (2) 52–57.

Guenther, D. A., A. J. Rosman. 1994. Differences between Compustat and CRSP SIC codes and related effects on research. *J. Accounting Econom.* **18** 115–128.

Guler, I., M. F. Guillén, J. M. MacPherson. 2000. Global competition, institutions, and the diffusion of organizational practices: The international spread of the ISO 9000 quality standards. *Admin. Sci. Quart.* **47**(2) 207–232.

Hair, J. F., R. E. Anderson, R. L. Tatham, W. C. Black. 1998. *Multivariate Data Analysis*, 5th ed. Prentice Hall, Upper Saddle River, NJ.

Heckman, J. J. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* **46**(4) 931–959.

Hendricks, K. B., V. R. Singhal. 1996. Quality awards and the market value of the firm: An empirical investigation. *Management Sci.* **42**(3) 415–436.

Hendricks, K. B., V. R. Singhal. 1997. Does implementing an effective TQM program actually improve operating performance? Empirical evidence from firms that have won quality awards. *Management Sci.* **43**(9) 1258–1274.

Hendricks, K. B., V. R. Singhal. 2001a. The long-run stock price performance of firms with effective TQM programs. *Management Sci.* **47**(3) 359–368.

Hendricks, K. B., V. R. Singhal. 2001b. Firm characteristics, total quality management, and financial performance. *J. Oper. Management* **19** 269–285.

Heras, I., G. P. M. Dick, M. Casadesús. 2002. ISO 9000 registration's impact on sales and profitability: A longitudinal analysis of performance before and after certification. *Internat. J. Quality Reliability Management* **19**(6) 774–791.

Hill, T. 1993. *Manufacturing Strategy: Text and Cases*. IRWIN.

Hsiao, C. 1993. *Analysis of Panel Data*. Cambridge University Press, Cambridge, UK.

ISO. 2003. *The ISO Survey of ISO 9000 and ISO 14000 Certificates.* (Twelfth cycle: Up to and Including 31 December 2002.) ISO Central Secretariat, http://www.iso.ch/.

Ittner, C. D., V. Nagar, M. V. Rajan. 2001. An empirical examination of quality-based learning models. *Management Sci.* **47**(4) 563–578.

King, A., M. Lenox. 2001. Lean and green? An empirical examination of the relationship between lean production and environmental performance. *Production Oper. Management* **10**(3) 244–256.

King, A., M. Lenox. 2002. Exploring the locus of profitable pollution reduction. *Management Sci.* **48**(2) 289–299.

Lewellen, W. G., S. G. Badrinath. 1997. On the measurement of Tobin's Q. *J. Financial Econom.* **44** 77–122.

Lie, E. 2001. Detecting abnormal operating performance: Revisited. *Financial Management* **30**(2) 77–91.

Lima, M. A. M., M. Resende, L. Hasenclever. 2000. Quality certification and performance of Brazilian firms: An empirical study. *Internat. J. Production Econom.* **66** 143–147.

Martínez-Costa, M., A. R. Martínez-Lorente. 2002. Effects of the ISO 9000 certification on the firm's performance: A vision from the market. Presented at *Production and Operations Management Conference*, San Francisco, CA.

Mendel, P. J. 2001. International standardization and global governance: The spread of quality and environmental management standards. A. Hoffman, M. Ventresca, eds. *Organizations, Policy, and the Natural Environment: Institutional and Strategic Perspectives.* Stanford University Press, Stanford, CA.

Mukherjee, A. S., M. A. Lapré, L. N. Van Wassenhove. 1998. Knowledge driven quality improvement. *Management Sci.* **44**(11, Part 2) S35–S49.

Naveh, E., A. Marcus. 2004. When does ISO 9000 quality assurance standard lead to performance improvements? Assimilation and going beyond. *IEEE Trans. Engrg. Management* **51**(3) 352–363.

Nicolau, J. L., R. Sellers. 2002. The stock market's reaction to quality certification: Empirical evidence from Spain. *Eur. J. Oper. Res.* **142**(3) 632–641.

Powell, T. C. 1995. Total quality management as competitive advantage: A review and empirical study. *Strategic Management J.* **16** 15–37.

Reimann, C. W., H. S. Hertz. 1994. Understanding the important differences between the Malcolm Baldrige National Quality Award and ISO 9000 Registration. *Production Oper. Management* **3**(3) 171–185.

Russo, M. V. 2001. Institutional change and theories of organizational strategy: ISO 14000 and environmental management in the electronics industry. Working paper, University of Oregon, Eugene, OR.

Simmons, B. L., M. A. White. 1999. The relationship between ISO 9000 and business performance: Does registration really matter? *J. Managerial Issues* **11**(3) 330–343.

Singels, J., G. Ruël, H. van de Water. 2001. ISO 9000 series: Certification and performance. *Internat. J. Quality Reliability Management* **18**(1) 62–75.

Smith, J. A., P. E. Todd. 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? *J. Econometrics* **125**(1–2) 305–353.

Terlaak, A., A. King. 2005a. The effect of certification with the ISO 9000 quality management standard: A signaling approach. *J. Econom. Behavior Organ.* Forthcoming.

Terlaak, A., A. A. King. 2005b. Choosing whom to follow: Profitability expectations and the role of smaller organizations in information-based adoption processes. Manuscript, University of Wisconsin-Madison, WI.

Terziovski, M., D. Power, A. S. Sohal. 2003. The longitudinal effects of the ISO 9000 certification process on business performance. *Eur. J. Oper. Res.* **146** 580–595.

Terziovski, M., D. Samson, D. Dow. 1997. The business value of quality management systems certification: Evidence from Australia and New Zealand. *J. Oper. Management* **15** 1–18.