Information Systems Research
Vol. 20, No. 2, June 2009, pp. 295–313
ISSN 1047-7047 | EISSN 1526-5536 | 09 | 2002 | 0295

informs

DOI 10.1287/isre.1080.0184

From Association to Causation via a Potential Outcomes Approach

Sunil Mithas
Robert H. Smith School of Business, University of Maryland, College Park, Maryland 20742, smithas@rhsmith.umd.edu
M. S. Krishnan
Ross School of Business, University of Michigan, Ann Arbor, Michigan 48109, mskrish@umich.edu

Despite the importance of causal analysis in building a valid knowledge base and in answering managerial questions, the issue of causality rarely receives the attention it deserves in information systems (IS) and management research that uses observational data. In this paper, we discuss a potential outcomes framework for estimating causal effects and illustrate the application of the framework in the context of a phenomenon that is also of substantive interest to IS researchers. We use a matching technique based on propensity scores to estimate the causal effect of an MBA on information technology (IT) professionals' salary in the United States. We demonstrate the utility of this counterfactual or potential outcomes–based framework in providing an estimate of the sensitivity of the estimated causal effects because of selection on unobservables. We also discuss issues related to the heterogeneity of treatment effects that typically do not receive as much attention in alternative methods of estimation, and show how the potential outcomes approach can provide several new insights into who benefits the most from the interventions and treatments that are likely to be of interest to IS researchers. We discuss the usefulness of the matching technique in IS and management research and provide directions to move from establishing association to assessing causation.

Key words: business value of IT; returns on MBA; causal analysis; propensity score; matching estimator; counterfactual approach; treatment effect heterogeneity; selection on unobservables; sensitivity analysis; MBA; IT professionals
History: Ritu Agarwal, Senior Editor; Sarv Devaraj, Associate Editor. This paper was received on January 4, 2005, and was with the authors $19\frac{1}{2}$ months for 4 revisions. Published online in *Articles in Advance* December 18, 2008.

1. Introduction

Practitioners, academic researchers, and philosophers have a significant interest in assessing causal relationships. Consider an information technology (IT) professional who wants to improve his or her career prospects. This professional poses the following question: Am I better off with an MBA than without one? In other words, does an MBA cause an IT professional's salary to increase? Top executives and senior IT managers routinely pose such questions and inquire whether they are better off with or without a certain IT or strategic intervention. Academic researchers also acknowledge that a relationship is not compelling if causality cannot be established (Boulding et al. 2005). Causality has also been an important concern in the history of philosophy, and it is one of the four important issues (along with generalizability, explanation, and prediction) underlying the understanding of theory (Gregor 2006).

Although researchers have debated the issues related to causality and though econometric literature proposes several approaches to estimate causal relationships, much of this literature has either been too philosophical or made assumptions that are often doubtful or not testable. For example, Aral et al. (2006, p. 1820) note that "determining causality is essential to understanding whether IT pays off . . . [A] definitive answer to this question has defied purely econometric solutions, such as instrumental variables, because good instruments generally do not exist." Fortunately, recent advances in statistics and econometrics and associated methodological innovations offer an opportunity to investigate causal questions in a rigorous and satisfactory manner. At the same time,

these advances allow researchers to assess the plausibility and sensitivity of the causal estimates.

In this paper, we discuss a potential outcomes approach that uses propensity scores, thus responding to Gregor's (2006, p. 635) call that "the issues of causality,... could be analyzed in greater depth with the aim of making argument about these issues more accessible" to information systems (IS) researchers. We also discuss issues related to heterogeneity of treatment effects that typically do not receive as much attention in alternative methods of estimation and show how the potential outcomes approach can provide several new insights into who benefits the most from interventions and treatments that are of interest to IS and management researchers. Finally, we discuss a way to quantify the sensitivity of the estimated causal effect.

The potential outcomes approach we describe has its origins in Neyman's (1990, originally published in 1923 in Polish) work, which introduced the notion of potential outcomes in the context of experimental data. Rubin (2005 and the references therein) extends Neyman's work to analyze causal effects in observational studies. This approach has received considerable attention in statistical (Dehejia and Wahba 1999, Holland 1986), philosophical (Glymour 1986), epidemiological (Little and Rubin 2000), sociological (Winship and Sobel 2004), and econometric (Dehejia and Wahba 2002; Heckman et al. 1997, 1998b) literature. Although recent work (e.g., Mithas et al. 2005, 2006) provides some glimpses of the promise of this approach in management research, there is a need for a more complete articulation of this method in the context of IS and management literature for broader usage.

Unlike the associational relationships that are based on explained variance (as in regression), explained covariance (as in structural equation models, such as LISREL or TETRAD[1]), prediction (as in partial least squares), or a comparison of performance at time t with some prior time $t-1$ (as in "before-and-after" study designs), the notion of causality in the counterfactual or potential outcomes framework views causal effect as a contrast or a comparison between two potential outcomes at a given time t corresponding to a treatment or intervention that can be applied to each unit. This way of assessing causality comes close to answering the counterfactual questions that IT professionals, executives, policymakers, and researchers face. While IT professionals ask whether they are better off investing in an MBA, executives contemplate a comparison between two scenarios: one in which a firm invests in an IT or some other strategic intervention and one in which the firm does not. Researchers also think of causal effects in terms of comparison of performance at a given time in a counterfactual sense (as opposed to a before-and-after sense) (Barney 2002, Porter 1987). For example, Barney (2002, p. 162) notes that a firm's resources and capabilities "are valuable if, and only if, they reduce a firm's costs or increase its revenues compared to what would have been the case if the firm did not possess those resources." Likewise, Porter (1987, pp. 45–46) notes, "Linking shareholder value quantitatively to diversification performance only works if you compare the shareholder value that is with the shareholder value that might have been without diversification."

We structure the rest of the article as follows: In the next section, we discuss the research context, the data, and a conventional econometric analysis as a baseline. Then, we discuss and illustrate the potential outcomes approach. Finally, we discuss of the advantages and challenges in implementing the propensity score approach, offer a comparison of the propensity score approach with other approaches, and suggest opportunities for using this approach in IS and management research.

2. The Research Context and Conventional Econometric Analysis

We illustrate the potential outcomes approach in the context of the following question: What is the causal effect of an MBA on an IT professional's salary? Unlike previous research that focuses on associational relationships linking human capital and institutional factors with IT professionals' compensation (e.g., Ang et al. 2002, Levina and Xin 2007, Mithas 2008, Mithas and Krishnan 2008), we focus on a narrow causal question involving a binary treatment (i.e., an MBA) to illustrate how the potential outcomes approach can provide new insights into causality that are not as easily captured in conventional regression-like approaches.

[1] For a discussion of TETRAD approach, see Lee et al. (1997).

Table 1 Variable Definition

Variable	Definition/operationalization
totcompcy	Total compensation of IT professionals by adding their base pay, bonus, and stock options.
Age	Respondent age measured in number of years.
edubach	Whether the respondent's highest educational degree is a bachelor's (1 = yes, 0 = no).
edumaster	Whether the respondent's highest educational degree is a master's other than an MBA (1 = yes, 0 = no).
edusomecoll	Whether the respondent's highest educational degree is some college education (1 = yes, 0 = no).
eduphd	Whether the respondent's highest educational degree is a Ph.D. (1 = yes, 0 = no).
edumba	Managerial competency of a respondent, measured by whether he or she has an MBA (1 = yes, 0 = no).
itexp	Technical competency, measured by a respondent's IT experience in number of years.
currcoexp	Tenure at the current firm measured in number of years.
Male	Indicates the gender of the respondent (male = 1, female = 0).
empno	Denotes organization size and is a bracketed variable that indicates a range for the number of employees in the respondent's firm (1 = fewer than 100, 2 = 101–1,000, 3 = 1,001–10,000, 4 = more than 10,000).
npg	Indicates a respondent's industry sector (1 for nonprofits and governmental organizations, 0 otherwise).
itind	Indicates the type of industry to which the respondent belongs (1 for IT vendors and service providers including telecommunications, 0 otherwise).
dotcom	Indicates type of firm (1 if the respondent works with a dot-com type of firm, 0 otherwise).
hrsperwkcy	The average number of hours per week put in by the respondent.
headhunterpm	Frequency of headhunter contact per month as a measure of ability.

We obtained archival data from the 2006 National Salary Survey conducted by *InformationWeek*, a leading and widely circulated IT publication in the United States. The salary survey covered more than 9,000 IT professionals and contains information about respondents' salaries, respondents' demographics, human capital–related variables, and institutional variables.[2] Table 1 provides the definition and construction for the variables we used in this research. Table 2 provides summary statistics for the independent variables across MBA and non-MBA IT professionals in our sample before matching. Note that MBA and non-MBA IT professionals differ in terms of their observed characteristics, and some of these differences are statistically significant.

[2] For more details on the *InformationWeek* salary surveys and data, refer to Mithas and Krishnan (2008).

Table 2 Characteristics of Treatment and Control Groups Before Matching

	Non-MBA[1]	MBA
N	9,108	675
itexp	14.89	15.87***
currcoexp	8.07	7.67
hrsperwkcy	47.29	48.92***
Age	42.22	43.17**
Male	0.84	0.83
empno	2.65	2.94***
npg	0.13***	0.09
itind	0.15	0.19***
dotcom	0.04	0.04
headhunterpm	0.24	0.32***

Notes. Significance levels for differences in means using *t*-tests on the larger of the two numbers across treatment and control units. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

[1] As regards the distribution of educational qualification among non-MBA IT Professionals, 49% have a bachelor's, 18% have a master's, 14% have some college education experience, 2% have a Ph.D. and remaining IT professionals have a high school diploma, associate degree, and any IT-related training after high school.

Although we focus here on methodological issues related to causality, we note an important threefold role of theory in the potential outcomes framework for causal inference. First, the causal approach we discuss uses the notion of "theory as explanation" to discuss why a certain intervention might be associated with a particular outcome (Gregor 2006). While theory, in the sense of explaining "why," provides a description of the likely intervening processes and mechanisms, unlike a structural equation modeling approach, in which researchers explicitly model these intervening variables, the goal in most empirical studies that use the potential outcomes approach is to calculate the total or reduced-form effect only (for a discussion of "direct" and "indirect" effects in potential outcomes approach, see Holland 1988 and Rubin 2004).

In our context, economic theories that focus on human capital endowments provide an explanation for why investments in an MBA can lead to higher compensation (see Gerhart and Rynes 2003 for a review of various compensation theories). In IS research, a wide body of academic and practitioner literature provides support for the effect of IT professionals' managerial competencies (e.g., those that can be acquired through an MBA) on career prospects. IT professionals with business knowledge play an important role in shaping the information

infrastructure of a firm to help firms become agile and differentiate among their competition (Ferratt et al. 2005, Josefek and Kauffman 2003, Ray et al. 2005, Sambamurthy et al. 2003, Smaltz et al. 2006). Bharadwaj (2000) argues that managerial competencies are important in coordinating the multifaceted activities associated with successful implementation of IT systems and for the integration of IT and business planning.

Thus far, we have discussed the "explanation" for or somewhat indirect role of theory in a potential outcomes approach. A second and more direct role of theory in the potential outcomes approach is to select the variables to match treatment and control units. To this end, theory provides guidance in identifying variables that might be correlated with assignment to the treatment and outcomes of interest. The identification and inclusion of all such variables in models, as suggested by theory, will likely satisfy the strong ignorability assumption that underpins the propensity score approach. In essence, this assumption states that conditional on observed covariates, assignment to a treatment group (e.g., MBA) is independent of potential outcomes (e.g., salary with an MBA and salary without an MBA) (Rosenbaum and Rubin 1983b). Consistent with this role of theory in the potential outcomes framework and following previous research (see Table A1 in the online supplement),[3] we identify variables that are likely to be correlated with the acquisition of an MBA and with outcome measures. Following human capital theory, we use human capital variables, such as IT experience and current firm experience. The neoclassical theory suggests that workers want to maximize their total utility from all components of a job, and that higher pay or a compensating wage differential should accompany relatively difficult, risky, or more responsible jobs. Therefore, we control for hours of work by an IT professional. Consistent with the sorting and incentive-based view of efficiency wage theories and adverse selection models in agency theory, which argue that higher wages help offset the difficulty of monitoring employees and measuring their performance and that high-ability employees avoid lower-paying employers, we control for the frequency of "headhunter" contact (as a proxy of "ability") in our models. Following the postinstitutional or new-institutional view, which suggests the important role of institutional forces (e.g., historical precedent, equity beliefs, ability to pay) in wage determination, we use industry and firm controls (IT industry, nonprofits, and government) in our models.

Third, in a comprehensive causal analysis that includes a sensitivity analysis, theory can help identify missing variables that affect both the treatment assignment and the outcome variable. In this role, theory provides a way to identify variables that may have been missed to assess sensitivity to the potential violation of the strong ignorability assumption. We use this theoretical aspect of in our sensitivity analysis, as we discuss in greater detail subsequently.

2.1. Conventional Econometric Approach

We specify a standard cross-sectional earnings model to include a dummy variable (Z) that indicates whether an individual has an MBA. Let X represent a vector of observed characteristics associated with the respondent, and let Y represent the respondent's annual salary. Consistent with previous research (DiNardo and Pischke 1997), we use the following log-linear specification to estimate our wage models:

$$\ln Y_i = \alpha Z_i + \beta X_i + \varepsilon_i, \tag{1}$$

where α and β are the parameters to be estimated and ε is the error term associated with observation i.

Column 1 of Table 3 reports the results of fitting Equation (1) by ordinary least squares (OLS). Consistent with previous research (Ang et al. 2002, DiNardo and Pischke 1997, Mithas and Krishnan 2008), we calculate returns to an MBA while controlling for several covariates related to human capital endowments and respondent demographics in the regression equation.[4] The MBA coefficient in Column 1 of Table 3 provides an assessment of the compensation difference between IT professionals with an MBA and

[3] Additional information is contained in an online appendix to this paper that is available on the *Information Systems Research* website (http://isr.pubs.informs.org/ecompanion.html).

[4] We do not control for job titles because an MBA may enable IT professionals to qualify for better paying jobs at higher levels; therefore, including dummies for job titles in wage regression will underestimate the true returns from an MBA (see Angrist and Krueger 1999).

Table 3 Parameter Estimates ($N = 9{,}783$)

	(1)	(2)	(3)
	Natural log of total compensation (2005 dollars)[1]	Total compensation (2005 dollars)[1]	Propensity score model for an MBA (logit model)
Treatment, MBA or not	0.385*** (0.000)	38,574.245*** (0.000)	
itexp	0.017*** (0.000)	1,612.850*** (0.000)	0.007 (0.331)
hrsperwkcy	0.013*** (0.000)	1,669.930*** (0.000)	0.024*** (0.000)
headhunterpm	0.095*** (0.000)	12,754.750*** (0.000)	0.178*** (0.009)
R-square	0.320	0.154	
Chi-square			115.11[2]

Notes. Robust *p* values are in parentheses. *Significant at 10%; **significant at 5%; ***significant at 1%.

[1]The models include an intercept, tenure at current firm, age, gender, firm size, dummies for other education variables such as bachelor's degrees, master's degrees, Ph.Ds., or some college education, dummy for IT Industry, nonprofit sector, and whether the firm is a dot-com or not, interaction between IT industry and IT experience.

[2]df = 12, log-likelihood = −2,398.34, AIC = 0.817, BIC = −81,428.66, classification accuracy = 93.01%.

the reference group of IT professionals without an MBA (this reference group includes IT professionals with a high school diploma, those with an associate degree, and those with any IT-related training after high school). Column 1 in Table 3 shows that IT professionals with an MBA earn 47% ($100 \times \exp[0.38] - 1$) more than the reference group.[5] Thus, the results of this study provide evidence for a positive association between an MBA and IT professionals' salaries.

2.2. Can We Attribute Causal Interpretation to the Regression Results?

Although the preceding econometric analysis provides support for a positive and statistically significant association between an MBA and IT professionals' salaries, the nature of observational data raises concerns about the causal interpretation of our findings. Even if we assume that we entered all the covariates in Equation (1) with their correct functional form, which is rarely known (see Achen 2005), regression parameters based on observed data on outcomes and treatment assignment do not sustain a causal interpretation without additional assumptions that are analogous to the strong ignorability assumption we mentioned previously (i.e., conditional on observed covariates, assignment to MBA is independent of potential outcomes) (Holland 2001, Pratt and Schlaifer 1988).

The fundamental problem in causal inference is that we do not observe an individual subject (i) in two possible states (0 and 1) with associated outcomes (Y_{i0} and Y_{i1}) (Holland 1986).[6] In our context, assessing the causal effect of a treatment such as an MBA is difficult because we cannot observe the compensation of IT professionals with an MBA if they had not acquired the MBA, and vice versa. A statistical solution to this missing data problem can be attempted by reformulating the problem at the population level by defining the average treatment effect for a randomly selected subject from the population (Holland 1986). This presupposes that treatment and control groups are similar in all respects on both observed and unobserved characteristics. As mentioned previously, the nature of data collection for observational studies often fails to meet the assumption of the random treatment of assignment.

Selection problems do not arise in classical experimental settings in which subjects are randomly selected to either a treatment or a control group. If we consider the acquisition of an MBA as a form of exogenous treatment, an experimental setting would enable us to make the treatment and control groups equal in every respect (in both observed and unobserved characteristics); we could then assess the treatment effect by comparing the mean outcomes of the treatment and control groups at a given time after the treatment is assigned. In contrast, nonexperimental settings do not allow random treatment assignment

[5] To provide an easy interpretation of our results in dollar terms, we estimate wage regression with total compensation as the dependent variable (Column 2 of Table 3). We find that IT professionals with an MBA earn approximately $38,574 more than IT professionals who have a high school diploma, an associate degree, or any IT-related training after high school. An alternative comparison group for IT professionals with an MBA may be those with a BS or BA and two additional years of experience. Such a comparison indicates that an MBA provides a $16,074 advantage.

[6] Barney (2002, p. 189) describes this fundamental problem of causal inference in testing tenets of the resource-based view as follows: "[A] firm cannot compare its performance with resources and capabilities to itself without these resources and capabilities."

and may be affected by both observed and unobserved characteristics of subjects.[7] Therefore, a direct comparison of mean outcomes in observational studies may overestimate or underestimate the true causal effect of the relevant intervention (i.e., an MBA). Increasing the sample size does not remedy the problems stemming from selection on observable or unobservable characteristics.

Selection bias because of correlation between the observed characteristics of a subject and the subject's treatment status can be addressed by using a matching technique based on propensity scores (as we describe in §3.1), which also enables us to investigate heterogeneous treatment effects based on propensity scores (§3.2). In contrast, selection bias stemming from correlation between unobserved variables and a firm's treatment status is a more difficult problem. However, the use of the matching technique and related conceptual developments in statistics enables us to conduct a sensitivity analysis to assess how severe the selection on unobservables must be to question the inferences based on selection on observables (Rosenbaum 1999, 2002, 1987; Rosenbaum and Rubin 1983a). We discuss the sensitivity analysis in §3.3. The appendix provides a summary of step-by-step directions for the analyses we present in the §3.

3. A Potential Outcomes–Based Propensity Score Approach to Assess the Causal Effect

3.1. The Causal Effect Assuming Selection on Observable Characteristics

The first step in implementing a propensity score approach is to clearly identify the treatment, the outcome of interest, and other covariates. Usually, the selection of covariates depends on prior theory and data availability. In this study, we define the MBA as our treatment of interest, salary as an outcome, and several variables along which respondents with an MBA differ from those without an MBA as covariates (see Tables 1 and 2).

The second step involves defining the causal estimand using the potential outcomes approach. It is possible to define alternative causal estimands, such as the causal effect of an MBA for someone we pick randomly from the population or the causal effect of an MBA for someone who does not have an MBA (for a detailed discussion of alternative estimands, see Heckman 2000). In this study, we are interested in knowing the causal effect of an MBA for those who actually obtained an MBA. In other words, we focus here on the "average treatment effect on treated."

The third step in causal analysis involves making assumptions related to the observed data and the potential outcomes. As we noted previously, potential outcomes are essentially the outcomes that each respondent is presumed to have regardless of whether he or she is in the treatment or control group. For example, under the potential outcomes approach, each respondent would have two potential outcomes (i.e., a salary if he or she has an MBA and a different salary if he or she does not have an MBA). The fundamental problem in causal inference is that we can observe only one of the treatment states (a respondent will either have an MBA or not) and the associated outcome for each respondent. In other words, we have a "missing data" problem. The only way to solve this problem is by making some assumptions; by making a strong ignorability assumption, we solve the missing data problem and the fundamental problem of causal inference (see Rosenbaum and Rubin 1983b). We assume that selection on unobservables is ignorable; that is, we assume that selection bias is due only to correlation between observed subject characteristics and a subject's treatment status. More formally, this is equivalent to assuming that $(Y_1, Y_0) \perp\!\!\!\perp Z \mid x$, where Z denotes treatment status; Y_1, Y_0 denote potential outcomes with and without treatment, respectively; and $\perp\!\!\!\perp$ denotes independence (Rosenbaum and Rubin 1983b).

How well this assumption is satisfied in a given study requires comparing the set of matching variables in a study with those in prior studies.[8]

[7] These situations also arise in the context of the business value of IT research at the firm level. For example, a treatment, such as the deployment of a particular IT system, may stem from self-selection by managers or mandates by the government (e.g., Y2K), buyers, and industry consortiums (e.g., electronic data interchange and radio frequency identification), and treatment and control firms may vary significantly.

[8] Table A1 compares the matching variables we used in this study with those in previous studies (additional information is contained in the appendix to this paper, available on the *Information Systems Research* website (http://isr.pubs.informs.org/ecompanion.html)).

Given that our study captures many of the important variables that have previously been considered relevant, perhaps we are not far off in assuming strong ignorability.[9] At the very least, by capturing some of the salient observable variables, we reduce bias that could arise from these variables, even if there is any doubt that this is a complete set of variables that might yield a "correct" estimate of the causal effect. Subsequently, we conduct a sensitivity analysis to provide an estimate of the extent to which our study may be vulnerable to what we may have left out of our propensity score model because of unavailability of data or uncertainty about whether a certain variable should belong in the model.

Table 2 compares the observed characteristics of IT professionals with an MBA with those of IT professionals who do not have an MBA. Table 2 shows that before matching, compared with non-MBA IT professionals, IT professionals with an MBA are older, have more IT experience, are more likely to work longer hours, work in larger firms, are less likely to work in nonprofit and government sectors, and were contacted more frequently by a headhunter.

The fourth step in causal analysis involves selection of an estimation method. We select the kernel matching estimator (Heckman et al. 1998b), as we discuss subsequently, though propensity score subclassification or other matching estimators can also be used. We describe propensity score subclassification in the context of treatment effect heterogeneity. Usually, these alternative methods provide similar estimates.

The fifth step involves estimation of the causal effect by calculating the propensity score using a logit model and the kernel matching estimator. We use a logit[10] model for the selection of MBA status and use these observed characteristics of a respondent as covariates in the selection equation (refer to Table 3, Column 3). The chi-square test in Table 3, Column 3, shows that the selection model is significant compared with a model with no explanatory variables. Thus, IT professionals with an MBA differ significantly from those who do not have an MBA with respect to observable characteristics. Thus, we reject the hypothesis of random assignment of MBA status among IT professionals. Favorable characteristics, such as the ability to put in longer hours, significantly increase the probability of acquiring an MBA because such professionals may benefit more from acquiring an MBA or may have better awareness of the potential benefits of an MBA given their profile and experience.

The propensity score, defined as $e_i(x_i) = \Pr(Z_i = 1 \mid x_i)$, is the conditional probability that a subject with $X = x$ will be in the treatment group (x_i is the observed vector of background variables, and Z_i denotes the treatment status for individual i). We calculated the propensity score using a logit model (see Table 3, Column 3). Hirano et al. (2003) show that the use of estimated propensity scores leads to more efficient estimation than the use of true propensity scores. Because the matching estimators do not identify the treatment effect outside the region of common support on the propensity score, we first calculate the range of support for both the MBA and the non-MBA groups. In our study, the propensity score support for the treatment (i.e., MBA) group is 0.022–0.26, and the propensity score support for the control (i.e., non-MBA) group is 0.018–0.40. Bias stemming from nonoverlapping support for $e(x)$ can be a large part of selection bias (Heckman et al. 1998a); however, this may not be an issue in this research because we lost only 16 subjects out of 9,783 because they were outside the common support. Of the 16 subjects (none of whom have an MBA) outside the common support, 15 have a propensity score of less than 0.022, and one has a propensity score of 0.40. Rubin (2005) argues that such subjects should be dropped from analysis because it is not possible to find an appropriate "clone" for them. Thus, we do not consider

[9] It is impossible to specify a perfectly complete model for a propensity score because of likely ignorance of or disagreement about all the variables that may affect a phenomenon. Even if a researcher knows about all the important variables, it may not be possible to collect data on all these variables in a given study because of time or resource constraints that researchers realistically face.

[10] Because the choice of a specific model (probit or logit) does not affect the estimated probability of selection, the use of a probit model would be equally appropriate. As Heckman (2005, p. 65) notes, the method of propensity score matching does not require "exogeneity of conditioning variables" in the propensity score model, because we assume strong ignorability, and by implication, we assume that variables not included in the propensity score model are uncorrelated with regressors in the propensity score model.

Table 4 Overall Treatment Effect on Treated Using Kernel Matching

	Treated	Controls	Difference
Total compensation in 1999 dollars			
Before matching	122,969	93,437	29,532
After matching*	122,969	93,815	29,154**

*Kernel matching using Gaussian Kernel. Of 9,783 IT employees, only 9,767 had common support, and 675 had an MBA. We lost 16 observations during matching that were off support.

**Average treatment effect on treated.

these 16 subjects that fell outside the common support region (propensity scores for these firms are either less than 0.02 or more than 0.26) in our subsequent analysis.

Traditional propensity score matching methods pair each treatment subject with a single control subject, such that pairs are chosen on the basis of similarity in the estimated probabilities of selection into the treatment. Following recent work in econometrics (Heckman et al. 1997, 1998b) that extended traditional pairwise matching methods to a kernel matching estimator that uses multiple control subjects in constructing each of the matched outcome, leading to a reduction in variance of the estimator, we computed the average treatment effect for subjects with an MBA by matching all MBA subjects within the common support with a weighted average of all non-MBA subjects with weights that are inversely proportional to the distance between the propensity scores of MBA and non-MBA subjects. We specified the Gaussian function for the kernel matching. After matching on the propensity score and thus adjusting for the observed characteristics, we find that the average MBA effect is \$29,154 (see Table 4).

3.2. Treatment Effect Heterogeneity

The sixth step in a comprehensive causal analysis involves assessing the treatment effect heterogeneity based on propensity score. In our context, the use of propensity scores enables us to answer the question, Do all IT professionals benefit equally if they acquire an MBA? Researchers who study the business value of IT have argued that treatment effects can be heterogeneous. For example, Lucas (1993, p. 367) notes,

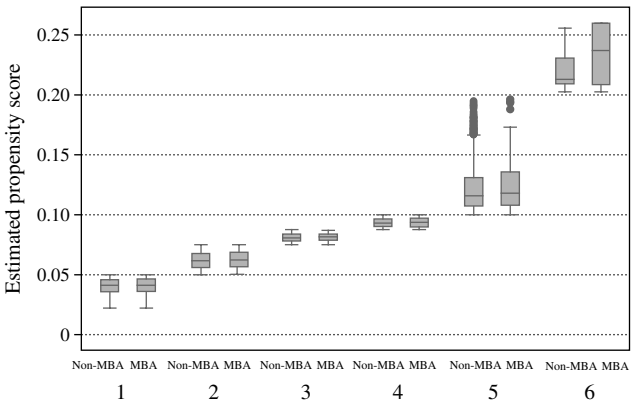
> A negative relationship between use and performance should not be immediately interpreted as an example of low business value from IT. It is quite possible that technology may help raise average performance by improving results for the lowest performing groups. There may be little market potential left for high performing companies and individuals.

Heterogeneity is even more plausible for the individual level phenomena (Xie and Wu 2005). Conventional regression approaches do not allow researchers to investigate heterogeneous treatment effects that vary depending on the propensity of being in the treatment group.

We used propensity score stratification to assess treatment effect heterogeneity. We formed subclasses (or strata) $S(x)$ on the basis of the estimated propensity scores (Rosenbaum and Rubin 1984) to reduce the dimensionality of observed covariates (i.e., X is reduced to $e_i[x_i]$, which is reduced to the subclasses $S[x]$). Rosenbaum and Rubin (1983b) show that conditioning on the univariate propensity score $e(x)$ (and, thus, $S[x]$) is equivalent to conditioning or matching on the entire multivariate X. In other words, subclassification based on the propensity score ensures that treatment and control units have similar values of the propensity score (thus achieving a balance on the multivariate X), enabling a "fair comparison" of treatment and control units within each subclass. Following Dehejia and Wahba (2002), we initially classified all observations in five equal-sized subclasses based on propensity scores. Then, we checked for any differences in propensity scores across treatment and control units in each stratum. If we found any significant differences, we subdivided the stratum until we obtained a similar distribution of propensity scores and covariates in each stratum. This resulted in six strata that achieved a propensity score and covariate balance across treatment and control units.

Figure 1 shows the distribution of propensity scores of treated and control subjects in each stratum. The matching technique tries to distribute the mean value of x equally between MBA and non-MBA subjects within each stratum, so the estimated treatment effect has no bias because of mean x. Table 5 shows the covariate balance and summary statistics across treatment and control units within each stratum after subclassification based on propensity scores and the reduction in bias achieved through matching on x. The mean on observed x in Table 5 is much closer

Figure 1 Propensity Score Distribution Across Strata



between treatment and control groups after matching based on propensity score stratification than before matching (see Table 2).

Table 6 shows the difference in compensation for MBA and non-MBA subjects after accounting for selection bias from correlation between observed variables (used in the selection equation) and the treatment variable in each stratum. Figure 2 shows a plot of the estimated average treatment effect within each stratum. This plot shows evidence for treatment effect heterogeneity across strata. The results suggest that the gains from an MBA are initially higher in Stratum 1, somewhat lower and flat in Strata 2 and 3, rise again in Strata 4 and 5, and are the highest in Stratum 5. Note also that Stratum 6 shows a negative effect of an MBA; we do not make much of this result because of the small sample size in this stratum. Broadly speaking, these results imply that subjects who were least likely to obtain an MBA (those in Stratum 1) benefit significantly from obtaining an MBA; that is, they experience higher gains than those who were more likely candidates for an MBA (e.g., those in Strata 2 and 3). Turning to the interpretation of the results in Strata 4 and 5, it appears that gains from an MBA are substantial even if one has a higher probability to pursue an MBA (perhaps because of demographic characteristics), though such candidates might also have higher opportunity costs of pursuing MBA because these characteristics may make them successful even if they do not pursue

Table 5 Characteristics of Treatment and Control Groups After Matching

	Stratum 1		Stratum 2		Stratum 3		Stratum 4		Stratum 5		Stratum 6	
	Non-MBA	MBA	Non-MBA	MBA	Non-MBA	MBA	Non-MBA	MBA	Non-MBA	MBA	Non-MBA	MBA
Panel A												
N	2,631	94	3,358	244	1,226	91	794	102	1,065	140	18	4
itexp	12.01	13.21	14.72	15.35	16.23	15.37	17.05	17.56	19.75	18.22	23.22	22.50
currcoexp	8.97	9.95	8.34	8.44	8.17	7.43	7.36	7.04	5.74	5.48	3.72	4.00
hrsperwkcy	43.82	43.90	46.75	47.03	48.29	48.33	49.96	48.73	54.33	55.63	66.67	70.75
Age	39.72	40.76	42.25	42.67	43.11	42.46	44.00	44.50	46.14	44.85	49.17	49.75
Male	0.85	0.83	0.85	0.86	0.84	0.85	0.83	0.86	0.84	0.80	0.83	0.75
empno	1.70	1.76	2.62	2.60	3.32	3.36	3.53	3.48	3.66	3.69	3.89	4.00
npg	0.27	0.35	0.10	0.07	0.04	0.05	0.03	0.01	0.01	0.00	0.00	0.00
itind	0.06	0.07	0.12	0.09	0.18	0.18	0.27	0.31	0.32	0.34	0.44	0.50
dotcom	0.03	0.02	0.04	0.04	0.04	0.04	0.03	0.03	0.05	0.06	0.06	0.00
headhunterpm	0.09	0.15	0.19	0.19	0.26	0.25	0.34	0.47	0.64	0.58	2.04	1.67
Panel B												
Distribution of educational qualification among non-MBA IT professionals[1]												
edubach	0.48		0.5		0.5		0.51		0.46		0.56	
edumaster	0.15		0.17		0.2		0.21		0.25		0.22	
edusomecoll	0.15		0.14		0.12		0.13		0.13		0.17	
eduphd	0.01		0.02		0.02		0.04		0.04		0	

Notes. Significance levels for differences in means using *t*-tests on the larger of the two numbers across treatment and control units. We also conducted Hotelling's *T*-squared test, a specialized form of MANOVA, for all the covariates and the corresponding *F* statistic was nonsignificant across all strata suggesting successful matching on all observed covariates. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

[1]The omitted category of educational qualification includes IT professionals with a high school diploma, associate degree, and any IT-related training after high school.

Table 6 Propensity Score Stratification and Treatment Effect Heterogeneity

Stratum	Non-MBAs (untreated subjects)	MBAs (subjects in treatment)	Salary of non-MBAs (untreated subjects)	Salary of MBAs (subjects in treatment)	MBA salary premium
1	2,631	94	$73,957	$95,576	21,619***
2	3,358	244	90,383	107,762	17,379***
3	1,226	91	102,495	116,226	13,731***
4	794	102	108,491	132,148	23,657***
5	1,065	140	127,290	163,169	35,880**
6	18	4	252,306	206,750	−45,556

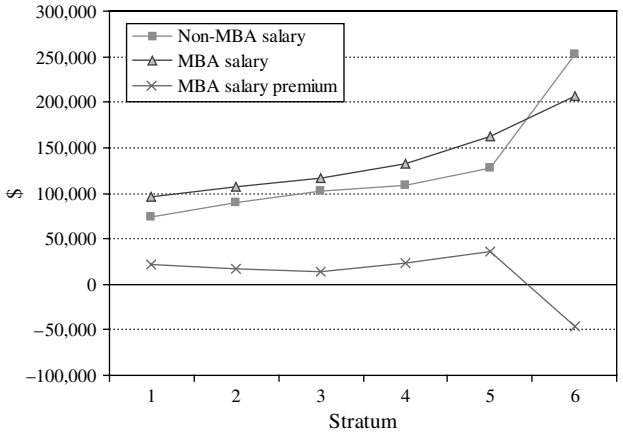
*Significant at 10%; **significant at 5%; ***significant at 1%.

MBA. This analysis provides a way to estimate treatment effect heterogeneity empirically in individual- and firm-level research, as called for by Lucas (1993) and other researchers (Xie and Wu 2005).

3.3. Sensitivity Analysis Assuming Selection on Unobservable Characteristics

The seventh and final step in a comprehensive causal analysis involves checking the sensitivity of the estimated causal effect to potential violations of the strong ignorability assumption (see Rosenbaum 1999). As we noted previously, our analysis has assumed that treatment and control groups are different because they differ on the observed variables in the data set, and that once we account for the observed variables by calculating the propensity score, the potential outcome is independent of the treatment assignment. If treatment and control groups differ on unobserved measures, a positive association between treatment status and performance outcome would not represent a causal effect. These unobserved variables can include unobserved characteristics of IT professionals, such as their motivation, risk attitude, innate intelligence, willingness to travel or migrate to other locations, cheerful disposition, and leadership abilities (Mithas and Lucas 2008).[11]

Figure 2 Treatment Effect Heterogeneity



Because it is not possible to estimate the magnitude of selection bias because of unobservable factors with nonexperimental data, we calculate the upper and lower bounds on the test statistics used to test the null hypothesis of the no-treatment effect for different values of unobserved selection bias (Rosenbaum 1999). Note that sensitivity analysis does not indicate whether biases are present or what their magnitudes are. It informs us only about the magnitude of biases that, if present, might alter inference. In other words, we evaluate how inferences about the treatment effect will be altered if an unobserved variable that also affects an outcome is likely to affect the probability of getting into the treatment group.

The treatment effect we estimated previously (in §3.1) is robust if the unobserved variable is equally present in the treatment and control groups. If some unobserved variable u (for expositional and interpretational simplicity, we assume that u is binary (see Rosenbaum 1987))[12] has an unequal presence in the treatment and control groups and if this u also affects the probability of selection into the treatment status, it is likely to affect our estimated treatment effect. If unobserved variables have no effect on the probability of getting into the treatment group or if there are no differences in unobserved variables across treatment and control groups, there is no unobserved selection bias. In this case, controlling for observed selection would produce an unbiased estimate of the treatment effect.

Table 7 provides the p values from Wilcoxon sign-rank tests for the average treatment effect on treated (i.e., those who have MBA) and sets different values of Γ (log odds of differential assignment

[11] We thank Rajeev Dehejia for pointing to motivation and risk attitudes as likely unobservables that might affect selection into an MBA status.

[12] Rosenbaum (1987) shows that for certain classes of unobserved covariates and permutation inferences, assuming a binary unobserved variable u provides the most conservative interpretation of sensitivity estimates.

Table 7 Sensitivity Analysis for the Effect of an MBA on Total Compensation*

Gamma (Γ)**	Significance level
1	0.000
2	0.001
2.25	0.03
2.5	0.19

*Sensitivity analysis is for the average treatment effect on treated (i.e., for IT professionals with an MBA).

**Log odds of differential assignment to treatment because unobserved factors.

to treatment because of unobserved factors). At each Γ, we calculate a hypothetical significance level "*p*-critical," which represents the bound on the significance level of the treatment effect in the case of endogenous self-selection into treatment status. In terms of interpretation, $\Gamma = 1.5$ implies that two subjects with exactly the same x vector differ in their odds of participating in the treatment by a factor of 1.5, or 50%. If changes in the neighborhood of $\Gamma = 1$ change the inference about the treatment effect, the estimated treatment effects are sensitive to unobserved selection bias. However, if a large value of Γ does not alter inferences about the treatment effect, the study is not sensitive to selection bias.

We find that the total compensation of IT professionals is sensitive to unobserved selection bias if we allow treatment and controls to differ by as much as 250% ($\Gamma = 2.5$) in terms of unobserved characteristics. This is a very large difference given that we have already adjusted for several key observed background characteristics typically used in prior research. If subjects with a high value of u are overrepresented in the treatment group, the estimated treatment effect of \$29,154 in total compensation overestimates the true treatment effect. If those who have a low value of u are overrepresented in the treatment group, the estimated treatment effect of \$29,154 underestimates the true treatment effect, and the true treatment effect is highly significant. It is important to realize that the analyses presented here represent the worst-case scenario. The confidence intervals for the effect of an MBA on total compensation will include zero only if (1) an unobserved variable caused the odds ratio of treatment assignment to differ between treatment and control groups by 2.5 and (2) the effect of this unobserved variable was so strong that it perfectly determined the effect attributed to the treatment for the treatment or control case in each pair of matched cases in the data. If the confounding variable has a strong effect on treatment assignment but only a weak effect on the outcome variable, the confidence interval for an outcome, such as total compensation, will not contain zero. Overall, the results of sensitivity analyses suggest that we will question the causal effects estimated previously (in §3.1) only if we believe that IT professionals with an MBA are 250% more likely than those without MBA to be endowed with any unobserved factor, such as better leadership, IQ, motivation, risk attitudes, or any other factor.

4. Discussion

The goal of this study is to show the usefulness of a propensity score matching technique for estimating the causal effect of a treatment or an intervention. We illustrated an application of this technique to estimate the causal effect of an MBA on the compensation of IT professionals during 2005 in the United States. We showed how use of this technique enables us to study heterogeneous treatment effects on the basis of the propensity of a subject to be in a treatment group, and we demonstrated how this technique enables us to quantify the likely bias in estimated causal effects because of selection on unobservables. We hope that by illustrating the use and advantages of a causal analysis, this research will encourage other researchers to apply this method to answer causal questions that arise in IS and management research at the firm and individual levels.

Although a counterfactual framework for assessing causality provides rich insights into and makes a researcher aware of the underlying assumptions and their implications before causal interpretation can be assigned to coefficients of regression or related models, we do not suggest that all research studies need to use this approach. There will continue to be a need for studies that enrich understanding by pointing to associations (Bharadwaj 2000; Mithas et al. 2007, 2008; Ramasubbu et al. 2008; Ray et al. 2004, 2005; Whitaker et al. 2007) or the nested nature of relationships (Ang et al. 2002, Mithas et al. 2006–07) that can subsequently be tested using the counterfactual framework we discussed herein. Likewise, there

is always a need for detailed case studies and historical accounts that help identify relevant variables to understand an unfolding phenomenon.

4.1. Advantages and Challenges in Using Propensity Score Approach

The propensity score approach has some notable advantages in the estimation of causal effects. First, this approach forces researchers to articulate their causal questions in terms of a comparison between two alternative states of the same unit, one with treatment and one without treatment. The use of such language forces researchers to be explicit about the treatment effect of substantive interest, the potential manipulability of the treatment (e.g., following this approach, a researcher would not view gender effects in the same way he or she would view returns to education), and the definition of the treated and control units. A related advantage of the propensity score approach is that it provides better visibility of the extent to which treatment and control groups are similar or different. This visibility enables researchers to ensure comparison of similar treatment and control units.

Second, the propensity score approach offers a solution to the curse of dimensionality that has impeded causal inquiry in previous research by forcing researchers to compare firms on only a few dimensions to avoid empty cells (as in matching based on covariates) without a fear of losing degrees of freedom (as in regression-based approaches, which typically specify a certain ratio of observations to parameters for statistical power and efficiency). In contrast, the propensity score approach allows matching on one covariate (propensity score), which in turn can be a function of multiple covariates, such as firm size, industry type, prior firm performance, and other sources of firm heterogeneity.

Third, the propensity score approach avoids functional form assumptions about the effects of covariates on the outcome variable, which are implicit in conventional regression analysis (Achen 2005, Dehejia and Wahba 1999, Rubin 1997). Although the propensity score approach also uses a parametric logit or probit specification in the first stage, this specification is relatively more flexible because an analyst does not need to worry about the potential loss of degrees of freedom if he or she were to use higher-order terms for covariates or wanted to include more covariates in the model.[13] More important, the approach is completely nonparametric in the second stage, when treated units are compared with their counterfactual control subjects; in this stage, treatment effects can be calculated at the individual or stratum level to assess the heterogeneity in treatment effects.

Finally, we show how the propensity score approach facilitates sensitivity analyses. Researchers often complain about the difficulty in articulating and justifying assumptions that are implicit in traditional econometric techniques (e.g., exclusion restrictions, as in the instrumental variables (IV) approach, or distribution of error terms, as in Heckman's selection models) to impute causal interpretation to the results of estimation (Aral et al. 2006, Briggs 2004). Because these assumptions are related to unobservable disturbances rather than to observable variables, researchers and managers find it difficult to interpret them substantively, thus making it difficult to understand and communicate findings (Angrist et al. 1996). In contrast, our sensitivity analysis allows for the quantification of a researcher's uncertainty; that is, researchers can quantify the severity of such assumptions and assess or debate the plausibility of the sensitivity of the causal effect to selection on unobservables (Angrist et al. 1996, Boulding et al. 2005, Mithas et al. 2005).

For all the reasons we have discussed, the propensity score method for assessing causality should be an attractive choice for researchers. However, this method has some challenges. First, a successful use of this method requires a relatively larger sample size so that matching is appropriately done and so that there is confidence in the matching quality. Therefore, researchers interested in causal effects of IT or strategic interventions with limited sample sizes may find this method somewhat restrictive. Second, this approach is currently most suitable and well developed for binary treatments. However, recent developments that have generalized the use of this technique for continuous treatments are encouraging (Imai and van Dyk 2004), and these developments may allow

[13] Unlike a one-step conventional regression, this specification can have any number of squared or interaction terms because the objective is to predict the probability of selection into treatment.

researchers to investigate the effect of continuous treatments (e.g., IT investments) on firm outcomes. Finally, because of the relatively nascent and evolving nature of this approach, the methodology has many flavors and is not as codified or standardized as regression-based approaches are in terms of statistical software or manuals. However, these limitations are being quickly overcome because researchers are beginning to share their software programs and make them available in public domain.[14]

4.2. Propensity Score Vis-à-Vis Other Approaches of Causal Inference

We briefly discuss how the propensity score approach is related to some other approaches for causal inference based on cross-sectional data in observational studies (for a detailed discussion, see Angrist and Krueger 1999, Winship and Morgan 1999, Winship and Sobel 2004). In particular, we compare and contrast the propensity score approach with regression, regression discontinuity (RD), the control function, and the IV approach.

Unlike regression-based approaches, which do not conceptualize how observed variables are related to the likelihood of assignment to the treatment group but instead rely on including all Xs in the model that are related to both the treatment and the outcome, RD and propensity score approaches attempt to control only for those observed variables that affect assignment to treatment to eliminate correlation between outcome and treatment based only on the observed Xs. An RD design relates an observed variable W (or a set of such variables) to the assignment to a treatment group (Campbell 1969) (for a detailed discussion and an empirical application of RD, see Hahn et al. 2001, DiNardo and Lee 2004, respectively). The basic idea in the RD approach is to find a W that is related to treatment assignment in a sharply discontinuous way and calculate the intercept or jump in the potential outcome locally at the point of treatment. The accuracy of treatment effect in this design depends on the relationship between W and Y in the absence of treatment over the range of W that receives the treatment. Note also that, unlike the propensity score approach in which we ensured that we had both treatment and control units over the relevant range of $e(X)$, in an RD approach, the opposite situation occurs and there are no values of W that contain both treatment and control units. In other words, the treatment effect in the RD approach depends on the ability to extrapolate accurately the values of Y for control units that do not receive treatment, and vice versa for the treated units.[15] Despite this difference, the RD design may be considered a special case of the propensity score approach because both share the same strategy of relating some observed variables to treatment assignment; whereas propensity score approach allows for estimation of the treatment effect over a range of W or $e(X)$ that has both treated and control units, RD only does local estimation in the vicinity of discontinuity of W or $e(X)$ and assignment to treatment changes.

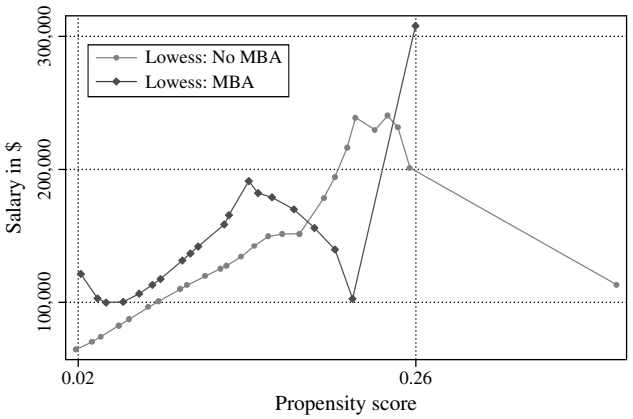
The propensity score method also shares a similarity with the control function (Heckman and Robb 1986, 1988) approach in that the propensity score, if it is included in a regression of the outcome on a treatment, makes it possible to estimate treatment effect consistently by making the treatment variable uncorrelated with the new error term (after the propensity score is included in the equation), assuming that a strong ignorability assumption holds (i.e., selection bias is based only on observables).[16] Heckman's (1978) work related to dummy endogenous variables, which allows for the correlation between error terms of selection and outcome equations (yet another example of a control function estimator), makes some additional assumptions about the relationships and distribution of error terms in the selection and outcome equations (i.e., a linearity assumption and a bivariate normality assumption,

[14] Additional information is contained in an online appendix to this paper that is available on the *Information Systems Research* website (http://isr.pubs.informs.org/ecompanion.html).

[15] An RD estimate of the treatment effect at a propensity score value of 0.02 in Figure 3 is $\$124,000 - \$65,932 = \$58{,}068$. This estimate based on an RD at a particular point is almost double the treatment effect estimated using propensity score matching, which uses a much wider range of propensity scores.

[16] We estimated the treatment effect using a control function approach by regressing total compensation on an MBA status and the propensity score. This procedure yielded a $21,274 treatment effect for an MBA.

Figure 3 The Regression Discontinuity Approach vs. Propensity Score Estimation*



*The bandwidth for lowess curves is 0.4. Whereas the RD approach calculates treatment effect only at propensity score of 0.02, the propensity score approach will use an interval of 0.02–0.26 in which both treated and control units are available.

respectively). In practice, although Heckman's estimator provides better estimates than OLS, the estimates can differ significantly from the experimental estimates, are highly sensitive to alternative specifications of selection equation (Briggs 2004), and do not allow for an assessment of the extent to which linearity and bivariate normality assumptions are satisfied (LaLonde 1986, Winship and Mare 1992). In contrast, the propensity score approach makes it possible to conduct a sensitivity analysis and numerically quantify the degree of uncertainty in a calculated treatment effect.

Finally, researchers often use an IV approach to solve endogeneity that arises when a regressor (t in our case, a treatment) is correlated with the error term (Greene 2000, Heckman 1997). This approach relies on finding an instrument variable R (or a set of such variables) that affects the assignment to a treatment but not to the outcome, an assumption (also known as the exclusion restriction [i.e., an instrument affects Y only through treatment and not otherwise]) that must be argued theoretically. Unlike a control function approach, which residualizes Y (outcome) and Z (treatment) with respect to some control function $e(X)$, such that residualized Z is no longer correlated with the resultant error term, an IV approach constructs a predicted Y and a predicted Z, in which the predicted Z is uncorrelated with the resultant error term by imposing an exclusion restriction (Winship and Morgan 1999). Note that Xs and Rs are conceptualized differently across the control function and the IV approaches. In the control function approaches (e.g., propensity score and endogenous dummy methods), Xs are so strongly correlated with both Z and Y that once these Xs, or a function thereof, are included in an outcome equation, it is assumed that the treatment indicator variable is no longer correlated with the remaining portion of the error term. In contrast, in an IV approach, one is interested in finding Rs that, by definition, are uncorrelated with the error term in the outcome equation.

Although the IV estimator is useful if valid instruments are available, in general, it is often difficult to justify the exclusion restrictions implicit in IV estimation (Bound et al. 1995; see Cascio and Lewis 2006 for an example). Moreover, this method assumes a constant treatment effect for all individuals and leads to large standard errors if instruments are weak and sample size is small. Angrist et al. (1996) discuss the IV estimator in the language of the potential outcomes framework and provide an insightful discussion about conditions under which IV estimates sustain a causal interpretation. Their work suggests that if an exclusion restriction and an additional monotonicity condition (the instrument either leaves the treatment unchanged or affects it in one direction only) for treatment assignment are satisfied, the conventional IV estimator is an estimate of a local average treatment effect (LATE) for compliers or defiers in a population. However, in general, it is difficult to identify these latent and unobserved groups, and the LATE estimate becomes specific to the choice of particular instrument.

On the whole, although propensity score approach offers a valuable perspective with which to conceptualize causal relationships and clarify the specific type of causal relationship, it should not be used as "a silver-bullet, black-box technique that can estimate treatment effect under all circumstances; ..." (Dehejia 2005, p. 363). Researchers must exercise their judgment on the applicability of this method and consider the type of research question, the research setting, the nature of the data, and the assumptions that they are willing to make (for a discussion, see Dehejia 2005; Heckman 2005; Smith and Todd 2005a, b).

Table 8 A Comparison of Approaches for Estimating Causal Effects

Approach	Key assumptions	Strengths	Limitations
Regression-based approach[1]	Error term is uncorrelated with all regressors	Easy to implement. Well-developed literature on mediating/moderating and nested models.	No clear distinction between treatment and covariates.
Propensity score approach	Strong ignorability (i.e., selection on observables only)	Estimates causal effect at a given time. Explicit consideration of all variables that relate to treatment assignment. Sensitivity analysis for violation of strong ignorability assumption.	Diagnostics for adequacy of propensity score model and methods for estimating mediating/moderating /nested effects are still in early stages. Requires large sample sizes.
RD approach	Strong ignorability	Allows estimation of treatment effect if treatment assignment changes discontinuously on the basis of some Z.	Requires making some assumptions and extrapolations for control units in the range of Z, where we do not have any control units, and vice versa.
Dummy endogenous variable approach	Error terms of selection and outcome equations are linearly related and bivariate normal	Allows error terms of selection and outcome equations to be correlated.	Linearity and bivariate normality assumptions are not testable. Both researchers and managers have difficulty in conceptualizing and understanding these assumptions.
IV approach	Exclusion restriction[2]	Allows estimation of causal effects when treatment variable is endogenous.	Exclusion restrictions are not testable and rarely justifiable. Large standard errors if sample sizes are small or instruments are weak. Assumes a constant treatment effect for all individuals.

[1]This approach sustains a causal interpretation only if an observability condition similar to the strong ignorability assumption is also assumed (Pratt and Schlaifer 1988).
[2]Instrumental variables affect outcome only through their effect on treatment variable.

Table 8 provides a summary of assumptions, strengths, limitations, and potential applications using alternative approaches for estimating causal effects that involve cross-sectional data. An appreciation of the underlying assumptions of these related methods of causal estimation and their substantive meaning in a particular context will lead to better conceptualization, dissemination, and consumption of IS and management research.

4.3. Opportunities for Using the Potential Outcomes Approach in IS and Management Research

We identify several areas for applying the potential outcomes view of causality in IS and management research. First, because the potential outcomes approach offers an alternative to randomization, which is almost impossible to achieve in the business value of IT research and similar questions in strategic management or marketing (for some quasi-experimental studies that seek to achieve randomization in field settings, see Banker et al. 1990), researchers will find the propensity score approach effective for answering research questions that focus on the estimation of causal effects of binary and sharp interventions on firm performance. These include initiatives such as implementation of enterprise or related IT systems (e.g., ERP, CRM, or SCM systems), the IT-enabled outsourcing of business processes, and many other marketing or strategic actions (Hitt et al. 2002, Mithas et al. 2005, Ray et al. 2005). Computing causal effects in a potential outcomes framework will provide a complementary perspective to the before-and-after perspective or a prediction-based perspective (based on a regression approach) currently in vogue.

Second, researchers can use the potential outcomes approach in settings that involve recursive path models to conceptualize the notion of causal effects in the potential outcomes framework (Holland 1988, Pearl 2000). Although we are not aware of any substantive empirical example of the potential outcomes approach in a structural equation modeling setting that uses cross-sectional data, researchers may find a variant of the potential outcomes approach—namely, marginal structural models (Robins et al. 2000)—useful if they have access to longitudinal data.

Third, causal effects have a clearer interpretation if all units can be considered potentially exposable to the treatment (Holland 1986). In line with this view, even if it is not appropriate to treat gender or industry effects in the same way as returns to an MBA, it is still possible to use the analytical apparatus that the potential outcomes approach offers in the sense of bias reduction.

Finally, an issue that IS researchers face is the amount of heterogeneity in organizational and individual units, which are likely to differ on dozens of dimensions. Therefore, we recommend the use of the potential outcomes approach to control for the known sources of heterogeneity for which data may be available because these sources of heterogeneity can be incorporated into the propensity score model without the fear of losing degrees of freedom or statistical power.

To conclude, this article articulates and elaborates the usefulness of a propensity score technique in the IS and management research by applying it to estimate the causal effect of an MBA on the compensation of IT professionals in the United States. We discuss issues related to the heterogeneity of treatment effects and show how the potential outcomes approach can provide several new insights into who benefits the most from interventions and treatments that are likely to be of interest to IS and management researchers. We also provide an estimate of the sensitivity of the estimated causal effects stemming from selection on unobservables. We hope that researchers will find the propensity score technique an attractive tool to answer their strategic and managerially relevant questions at firm and individual levels and to move from establishing association to understanding causation.

Acknowledgments

The authors thank the senior editor, the associate editor, and the anonymous Information Systems Research (ISR) reviewers for helpful comments on previous versions of this paper, as well as Daniel Almirall, Hemant Bhargava, Derek Briggs, Rajeev Dehejia, Rajiv Dewan, Chris Forman, Ben Hansen, Robert J. Kauffman, Hank Lucas, Barrie Nault, Steve Raudenbush, Edward Rothman, Donald Rubin, Paul Tallon, D. J. Wu, and Yu Xie; participants of the Winter 2004 Research Seminar on Causal Inference in the Social Sciences at the University of Michigan; and anonymous reviewers of our related substantive and methodology work for discussions and comments. A previous version of this paper using a different data set was presented at the INFORMS Conference on Information Systems and Technology (CIST 2004) in Denver, Colorado, where it won the best doctoral student paper award. The authors also thank *InformationWeek* for providing the necessary data and Srinivas Kudaravalli and Eli Dragolov for research assistance. Financial support for this study was provided in part by the Robert H. Smith School of Business at University of Maryland and R. Michael and Mary Kay Hallman fellowship at the Ross School of Business.

Appendix. Seven Steps to Implement a Propensity Score Approach

	Description
Step 1	Identify the treatment, the outcome of interest and other covariates. For example, in this study, treatment was an MBA, outcome was salary, and covariates were several variables along which respondents with MBA differ from those without an MBA. Usually, the selection of covariates depends on prior theory and data availability (see Tables 1 and 2, also Table A1).[a]
Step 2	Define the causal estimand using potential outcomes approach. For example, in this study, we calculated average treatment effect on treated because we were interested in knowing the causal effect of an MBA for those who actually obtained an MBA. One can define alternative causal estimands such as the causal effect of MBA for someone whom we pick randomly from the population or causal effect of MBA for someone who does not have an MBA (see Heckman 2000).
Step 3	Make assumptions that relate the observed data and the potential outcomes. Potential outcomes are essentially the outcomes that each respondent is presumed to have irrespective of whether she is in the treatment or control group. For example, under potential outcomes approach, each respondent would have two potential outcomes (i.e., a salary if she has an MBA and a different salary if she does not have an MBA). The fundamental problem in causal inference is that we can observe only one of the treatment states (a respondent will either have MBA or not) and the associated outcome for each respondent. In other words, we have a "missing data" problem here. The only way to solve this problem is by making some assumptions and by making a strong ignorability assumption we solve the "missing data" problem and the fundamental problem of causal inference (see Rosenbaum and Rubin 1983b).

Appendix. (Continued)

	Description
Step 4	Select an estimation method. We selected the kernel matching estimator in this paper (Heckman et al. 1998b), although one can also use propensity score subclassification or other matching estimators. We describe propensity score subclassification in the context of treatment effect heterogeneity.
Step 5	Estimate the causal effect by calculating the propensity score using a logit model and using the kernel matching estimator (see Tables 3 and 4).
Step 6	Assess the treatment effect heterogeneity based on propensity score. We formed six strata based on propensity scores and assessed the propensity score and covariate balance in each strata and then calculated the treatment effect for each strata separately (see Dehejia and Wahba 1999, 2002). See Figure 1 and Table 5 that provide an assessment of the success of the matching procedure using propensity score subclassification and Table 6 that shows estimated treatment effects in each stratum.
Step 7	Check the sensitivity of the estimated causal effect to potential violations of the strong ignorability assumption (see Rosenbaum 1999). See also Table 7.

[a]An online appendix to this paper is available on the *Information Systems Research* website (http://isr.pubs.informs.org/ecompanion.html).

References

Achen, C. 2005. Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management Peace Sci.* **22**(4) 327–339.

Ang, S., S. A. Slaughter, K. Y. Ng. 2002. Human capital and institutional determinants of information technology compensation: Modeling multilevel and cross-level interactions. *Management Sci.* **48**(11) 1427–1445.

Angrist, J. D., A. B. Krueger. 1999. Empirical strategies in labor economics. O. Ashenfelter, D. Card, eds. *Handbook of Labor Economics*. Elsevier, Amsterdam, 1277–1366.

Angrist, J. D., G. W. Imbens, D. B. Rubin. 1996. Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* **91**(434) 444–455.

Aral, S., E. Brynjolfsson, D. J. Wu. 2006. Which came first, IT or productivity? The virtuous cycle of investment & use in enterprise systems. *Proc. 27th Internat. Conf. Inform. Systems*, AIS, Milwaukee, WI, 1819–1840.

Banker, R., R. Kauffman, R. Morey. 1990. Measuring gains in operational efficiency from information technology: A case study of the positran deployment at Hardee's Inc. *J. Management Inform. Systems* **7**(2) 29–54.

Barney, J. B. 2002. *Gaining and Sustaining Competitive Advantage*, 2nd ed. Prentice Hall, Upper Saddle River, New Jersey.

Bharadwaj, A. 2000. A resource-based perspective on information technology capability and firm performance: An empirical investigation. *MIS Quart.* **24**(1) 169–196.

Boulding, W., R. Staelin, M. Ehret, W. J. Johnston. 2005. A customer relationship management roadmap: What is known, potential pitfalls, and where to go. *J. Marketing* **69**(4) 155–166.

Bound, J., D. A. Jaeger, R. M. Baker. 1995. Problems with instrumental variables estimation when the correlation between the instruments and endogenous explanatory variables is weak. *J. Amer. Statist. Assoc.* **90**(430) 443–450.

Briggs, D. C. 2004. Causal inference and the Heckman model. *J. Educational Behavioral Statist.* **29**(4) 397–420.

Campbell, D. T. 1969. Reforms as experiments. *Amer. Psych.* **24**(4) 409–429.

Cascio, E. U., E. G. Lewis. 2006. Schooling and the armed forces qualifying test: Evidence from school-entry laws. *J. Human Resources* **41**(2) 294–318.

Dehejia, R. 2005. Practical propensity score matching: A reply to smith and Todd. *J. Econometrics* **125**(1–2) 355–364.

Dehejia, R. H., S. Wahba. 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *J. Amer. Statist. Assoc.* **94**(448) 1053–1062.

Dehejia, R. H., S. Wahba. 2002. Propensity score-matching methods for nonexperimental causal studies. *Rev. Econom. Statist.* **84**(1) 151–161.

DiNardo, J. E., D. S. Lee. 2004. Economic impacts of new unionization on private sector employers: 1984–2001. *Quart. J. Econom.* **119**(4) 1383–1441.

DiNardo, J. E., J. S. Pischke. 1997. The returns to computer use revisited: Have pencils changed the wage structure too? *Quart. J. Econom.* **112**(1) 291–303.

Ferratt, T. W., R. Agarwal, C. V. Brown, J. E. Moore. 2005. IT human resource management configurations and IT turnover: Theoretical synthesis and empirical analysis. *Inform. Systems Res.* **16**(3) 237–255.

Gerhart, B., S. L. Rynes. 2003. *Compensation: Theory, Evidence, and Strategic Implications*. Sage Publications, Thousand Oaks, CA.

Glymour, C. 1986. Comment: Statistics and metaphysics. *J. Amer. Statist. Assoc.* **81**(396) 964–966.

Greene, W. H. 2000. *Econometric Analysis*, 4th ed. Prentice Hall, Upper Saddle River, NJ.

Gregor, S. 2006. The nature of theory in information systems. *MIS Quart.* **30**(3) 611–642.

Hahn, J., W. v. d. Klauww, P. Todd. 2001. Identification of treatment effects by regression discontinuity design. *Econometrica* **69**(1) 201–209.

Heckman, J. J. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* **46**(4) 931–961.

Heckman, J. J. 1997. Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *J. Human Resources* **32**(3) 441–462.

Heckman, J. 2000. Causal parameters and policy analysis in economics: A twentieth century retrospective. *Quart. J. Econom.* **115**(1) 45–97.

Heckman, J. J. 2005. The scientific model of causality. *Sociol. Methodology* **35**(1) 1–150.

Heckman, J. J., R. Robb. 1986. Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. H. Wainer, ed. *Drawing Inferences from Self-Selected Samples*. Springer-Verlag, New York, 63–113.

Heckman, J. J., R. Robb. 1988. The value of longitudinal data for solving the problem of selection bias in evaluating the impact of treatment on outcomes. G. Duncan, G. Kalton, eds. *Panel Surveys*. John Wiley, New York, 512–538.

Heckman, J. J., H. Ichimura, P. Todd. 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Rev. Econom. Stud.* **64**(4) 605–654.

Heckman, J. J., H. Ichimura, J. Smith, P. Todd. 1998a. Characterizing selection bias using experimental data. *Econometrica* **66**(5) 1017–1098.

Heckman, J. J., H. Ichimura, P. Todd. 1998b. Matching as an econometric evaluation estimator. *Rev. Econom. Stud.* **65**(2) 261–294.

Hirano, K., G. W. Imbens, G. Ridder. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71**(4) 1161–1189.

Hitt, L. M., D. J. Wu, X. Zhou. 2002. Investments in enterprise resource planning: Business impact and productivity measures. *J. Management Inform. Systems* **19**(1) 71–98.

Holland, P. W. 1986. Statistics and causal inference (with discussion). *J. Amer. Statist. Assoc.* **81**(396) 945–970.

Holland, P. W. 1988. Causal inference, path analysis, and recursive structural equation models. *Sociol. Methodology* **18**(1) 449–484.

Holland, P. W. 2001. The causal interpretation of regression coefficients. M. C. Galavotti, P. Suppes, D. Costantini, eds. *Stochastic Causality*. CSLI Publications, Stanford, CA, 173–187.

Imai, K., D. A. van Dyk. 2004. Causal inference with generalized treatment regimes: Generalizing the propensity score. *J. Amer. Statist. Assoc.* **99**(467) 854–866.

Josefek, R. A. J., R. J. Kauffman. 2003. Nearing the threshold: An economics approach to presure on information systems professionals to separate from their employer. *J. Management Inform. Systems* **20**(1) 87–122.

LaLonde, R. J. 1986. Evaluating the econometric evaluations of training programs with experimental data. *Amer. Econom. Rev.* **76**(4) 604–620.

Lee, B., A. Barua, A. B. Whinston. 1997. Discovery and representation of causal relationships in MIS research: A methodological framework. *MIS Quart.* **21**(1) 109–136.

Levina, N., M. Xin. 2007. Comparing IT worker's compensation across country contexts: Demographic, human capital, and institutional factors. *Inform. System Res.* **18**(2) 193–210.

Little, R. J., D. Rubin. 2000. Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Ann. Rev. Public Health* **21** 121–145.

Lucas, H. C. 1993. The business value of information technology: A historical perspective and thoughts for future research. R. D. Banker, R. J. Kauffman, M. A. Mahmood, eds. *Strategic Information Technology Management*: *Perspectives on Organizational Growth and Competitive Advantage*. Idea Group Publishing, Harrisburg, PA, 359–374.

Mithas, S. 2008. Are emerging markets different from developed markets? Human capital, sorting and segmentation in compensation of information technology professionals. *Proc. 29th Internat. Conf. Inform. Systems*, Association for Inforamtion Systems, Paris, France.

Mithas, S., M. S. Krishnan. 2008. Human capital and institutional effects in the compensation of information technology professionals in the United States. *Management Sci.* **54**(3) 415–428.

Mithas, S., H. C. Lucas. 2008. Does high-skill immigration make everyone better off? United States' visa policies and compensation of American and foreign information technology professionals. Working paper, Robert H. Smith School of Business, University of Maryland, College Park.

Mithas, S., D. Almirall, M. S. Krishnan. 2006. Do CRM systems cause one-to-one marketing effectiveness? *Statist. Sci.* **21**(2) 223–233.

Mithas, S., J. L. Jones, W. Mitchell. 2008. Buyer intention to use Internet-enabled reverse auctions? The role of asset specificity, product specialization, and non-contractibility. *MIS Quart.* **32**(4) 705–724.

Mithas, S., M. S. Krishnan, C. Fornell. 2005. Why do customer relationship management applications affect customer satisfaction? *J. Marketing* **69**(4) 201–209.

Mithas, S., N. Ramasubbu, M. S. Krishnan, C. Fornell. 2006–07. Designing websites for customer loyalty: A multilevel analysis. *J. Management Inform. Systems* **23**(3) 97–127.

Mithas, S., A. R. Tafti, I. R. Bardhan, J. M. Goh. 2007. Resolving the profitability paradox of information technology: Mechanisms and empirical evidence (available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1000732), Working paper, Robert H. Smith School of Business, University of Maryland, College Park.

Neyman, J. S., D. M. Dabrowska, T. P. Speed. 1990. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* **5**(4) 465–472.

Pearl, J. 2000. *Causality*: *Models, Reasoning, and Inference*. Cambride University Press, Cambridge, UK.

Porter, M. E. 1987. From competitive advantage to corporate strategy. *Harvard Bus. Rev.* (May–June) 43–59.

Pratt, J. W., R. Schlaifer. 1988. On the interpretation and observation of laws. *J. Econometrics* **39**(1–2) 23–52.

Ramasubbu, N., S. Mithas, M. S. Krishnan, C. F. Kemerer. 2008. Work dispersion, process-based learning, and offshore software development performance. *MIS Quart.* **32**(2) 437–458.

Ray, G., J. B. Barney, W. A. Muhanna. 2004. Capabilities, business processes, and competitive advantage: Choosing the dependent variable in empirical tests of the resource-based view. *Strategic Management J.* **25**(1) 23–37.

Ray, G., W. A. Muhanna, J. B. Barney. 2005. Information technology and the performance of the customer service process: A resource-based analysis. *MIS Quart.* **29**(4) 625–652.

Robins, J. M., M. A. Hernan, B. Brumback. 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology* **11** 550–560.

Rosenbaum, P. 1999. Choice as an alternative to control in observational studies. *Statist. Sci.* **14**(3) 259–304.

Rosenbaum, P. 2002. *Observational Studies*, 2nd ed. Springer, New York.

Rosenbaum, P. R. 1987. Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika* **74**(1) 13–26.

Rosenbaum, P. R., D. B. Rubin. 1983a. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J. Roy. Statist. Soc. Ser. B* **45**(2) 212–218.

Rosenbaum, P. R., D. B. Rubin. 1983b. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**(1) 41–55.

Rosenbaum, P. R., D. B. Rubin. 1984. Reducing bias in observational studies using subclassification on the propensity score. *J. Amer. Statist. Assoc.* **79**(387) 516–524.

Rubin, D. B. 1997. Estimating causal effects from large data sets using propensity scores. *Ann. Internal Medicine* **127**(8, Part 2) 757–763.

Rubin, D. B. 2004. Direct and indirect causal effects via potential outcomes. *Scandinavian J. Statist.* **31**(2) 161–170.

Rubin, D. B. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.* **100**(469) 322–331.

Sambamurthy, V., A. Bharadwaj, V. Grover. 2003. Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms. *MIS Quart.* **27**(2) 237–263.

Smaltz, D., V. Sambamurthy, R. Agarwal. 2006. The antecedents of CIO role effectiveness in organizations: An empirical study in the healthcare sector. *IEEE Trans. Engrg. Management* **53**(2) 207–222.

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

Smith, J. A., P. E. Todd. 2005b. Rejoinder. *J. Econometrics* **125**(1–2) 365–375.

Whitaker, J., S. Mithas, M. S. Krishnan. 2007. A field study of RFID deployment and return expectations. *Production Oper. Management* **16**(5) 599–612.

Winship, C., R. Mare. 1992. Models for sample selection bias. *Ann. Rev. Sociol.* **18** 327–350.

Winship, C., S. L. Morgan. 1999. The estimation of causal effects from observational data. *Ann. Rev. Sociol.* **25** 659–706.

Winship, C., M. Sobel. 2004. Causal inference in sociological studies. M. Hardy, A. Bryman, eds. *Handbook of Data Analysis*. Sage, London, 481–503.

Xie, Y., X. Wu. 2005. Market premium, social process, and statisticism. *Amer. Sociol. Rev.* **70**(5) 865–870.