Selection bias and econometric remedies in accounting and finance research

Jennifer Wu Tucker
Fisher School of Accounting
Warrington College of Business Administration
University of Florida
(352) 273-0214 (Office)
(352)392-7962 (Fax)
E-mail: jenny.tucker@warrington.ufl.edu

January 2011

Forthcoming in *Journal of Accounting Literature*

I thank Farid AitSahlia, Joel Demski, Mike Donohoe, Mark Flannery, Bill Greene, Marcus Kirk, Tom Omer, Doug Schroeder, and Senyo Tse for discussion and comments. I thank the Luciano Prida, Sr. Term Professorship Foundation for financial support. I thank Wayne Losano for editorial assistance. All errors are mine.

Selection bias and econometric remedies in accounting and finance research

ABSTRACT

While managers' accounting and financial decisions are, for many, fascinating topics, selection bias poses a serious challenge to researchers estimating the decisions' effects using non-experimental data. Selection bias potentially occurs because managers' decisions are non-random and the outcomes of choices not made are never observable. "Selection bias due to observables" arises from sample differences that researchers can observe but fail to control. "Selection bias due to unobservables" arises from the unobservable and thus uncontrolled sample differences that affect managers' decisions and their consequences. In this article I review two econometric tools developed to mitigate these biases—the propensity score matching (PSM) method to mitigate selection bias due to observables and the Heckman inverse-Mills-ratio (IMR) method to address selection bias due to unobservables—and discuss their applications in accounting and finance research. The article has four takeaways. First, researchers should select the correct method to alleviate potential selection bias: the PSM method mitigates selection bias due to observables, but does not alleviate selection bias due to unobservables. Second, in applying PSM researchers are advised to restrict their inferences to firms whose characteristics can be found in both the sample and control groups. Third, the IMR method, though popular, is limited to situations in which the choices are binary, the outcomes of choices are modeled in a linear regression, and the unobservables in the choice and outcome models follow multivariate normal distribution. Researchers can overcome these constraints by using full information maximum likelihood estimation. Last, when the IMR method is used, special attention should be paid to the formulas in calculating IMRs. The article also calls for researchers' attention to other approaches to evaluating the effects of managers' decisions.

Keywords: selection bias, propensity score matching, inverse mills ratio, Heckman model

1. Introduction

Many key decisions made at a firm can be categorized as "choices." The decisions range from operating, investing, and financing to financial reporting, voluntary disclosure, and executive compensation. For example, a retailer may adopt a just-in-time inventory system or a traditional one. Firms increase or decrease research-and-development expenditures, open new stores or close existing stores, and increase hiring or lay off employees. Managers may issue debts or equity to raise capital and further decide what particular types of security to issue. Firms may distribute dividends or repurchase stocks to return investments to investors. Managers may manipulate reported earnings. Anticipating an earnings shortfall, some managers warn investors but others do not. Firms may hire compensation consultants for CEO pay. These decisions fascinate accounting and finance researchers, who are interested in evaluating their consequences.

A serious challenge for researchers, however, is that, for a given firm, researchers observe only the outcome of the choice made but not the outcomes of choices not made. Therefore, researchers are unable to compare the outcome difference of choices for a given firm to evaluate the effects of its decision. To overcome this problem, researchers often attempt to select a control firm that is identical, except for the decision choice, to the firm that has made the choice of interest. This task is readily accomplished in controlled experiments, where subjects can be randomly assigned to treatment (i.e., the choice of interest) vs. non-treatment (i.e., the alternative choice) so that researchers can make inferences about the average effect of treatment. The task is problematic outside the realm of controlled experiments, where firms are heterogeneous and corporate decisions are observed ex post. What makes evaluations more challenging is that researchers cannot

observe all the information that managers and investors use in decision making. In other words, researchers use smaller sets of information to evaluate managers' decisions than the information sets used by managers and investors. It is crucial for researchers to account for the observable and unobservable differences between a selected control firm and the "ideal" control firm (that is, the sample firm itself) in evaluating treatment effects. Absent controls for these differences, selection bias, which is one form of endogeneity problem, can lead to inappropriate inferences about treatment effects. Examples of observable differences are firm size and growth. "Selection bias due to observables" results from a failure to control for differences researchers can observe. Unobservable differences arise because researchers are restricted to smaller information sets than managers and market participants. Examples of unobservables are information revealed during a financial audit that are known to some market participants or other information that is publically disclosed by the company but is too costly for researchers to collect.[1] "Selection bias due to unobservables" results from a failure to control for the effect of differences researchers cannot observe.

Both types of bias have been serious concerns for labor economists, who evaluate a variety of training and welfare programs. Statistical and econometric tools have been developed to address them. Matching a participant with a non-participant with similar (of course, observable) characteristics was a common tool used in the 1980s to mitigate selection bias due to observables [Heckman, Ichimura, and Todd 1998]. Matching on covariates (e.g., firm characteristics) is ideal when the number of characteristics over which participants and non-participants differ is limited and the variables are categorical.

[1] Throughout the article, the term "unobservables" means the factors affecting both treatment selection and treatment outcome. If an unobservable factor affects either process but not both, it does not cause an estimation bias.

Matching, however, is difficult or infeasible when the number of characteristics to match is large and the sample size is limited. Rosenbaum and Rubin [1983] propose matching by *a function of covariates* rather than by each covariate. The function they choose is the probability of an individual being selected into the program (and thus to be treated). This matching method is referred to as "propensity score matching" (PSM). PSM has been widely used and discussed in numerous disciplines, including statistics, economics, and medicine, in the past three decades and has been increasingly applied in accounting and finance research in the past few years. For example, *Journal of Accounting Economics* had no publications using PSM before 2010, but two publications in 2010; *Journal of Financial Economics* published its first two articles that use PSM in 2004 and three articles during 2006-2009, but published or accepted for publication 10 articles alone in 2010.

The concern of selection bias due to unobservables was first thoroughly addressed by Lee [1978] and Heckman [1979]. Heckman proposes a two-stage approach to evaluating programs for which the treatment choices are binary and the program outcomes depend on a linear combination of observable and unobservable factors. His approach is to estimate the choice model in the first stage and add a bias correction term in the second-stage regression. After further restricting unobservables to multivariate normal distributions, he derives the bias correction variable in the form of inverse Mills ratio (IMR), that is, a ratio of standard normal p.d.f. over standard normal c.d.f. (or 1 minus the c.d.f.) The IMR method is easy to implement and requires little computing power, and thus has become immensely popular. The method has been increasingly used in accounting and finance research in recent years. For example, *Journal of Accounting Research* published only one study that uses the IMR method by 1997, but nine studies since 2003. *Journal of*

Accounting and Economics published only one study with the IMR method by 1999, but 15 studies since 2005. The use of IMR in *Journal of Financial Economics* also surged in 2010, following a steady pace of applications since 1995.

The growing interest in PSM and IMR in accounting and finance research warrants a survey discussing the conditions for each method and the confusion and perhaps mistakes in their applications. In this article I share my observations regarding these issues. The PSM method requires "conditional independence," which means that the selection or self-selection of participant (treated firm) vs. non-participant (untreated firm) can all be explained by observable factors. The estimated treatment effect using PSM can only be generalized to "common support," meaning the portion of the population that can meaningfully decide whether to participate (unless all observations are used with weights that increase with the closeness of their match with the treated firm—a approach known as "kernel weighting"). The IMR method, on the other hand, deals with selection on unobservable factors. Because IMRs are derived from truncated binormal distributions, it is only appropriate if the first-stage choice decision is modeled in probit, and the second-stage outcome is modeled in a linear regression, and if the unobservables in the two stages are binormally distributed.[2] When these conditions are not met, adding IMR to the second stage does not correct the selection bias that researchers intend to correct.[3]

Despite the popularity of PSM and IMR, there is confusion in accounting and finance research about the appropriate use of each. One misconception is that PSM can address

[2] Wooldridge [2002, p.562-563] states that the binormal distribution assumption can be relaxed for the second-stage error term if its mean conditional on the first-stage error term (which has to be normal) is linear. However, except for binormal distributions, few bivariate distributions satisfy this condition.

[3] In this article I assume that all relevant observable variables are identified and reliably measured and that the form of their relations is correctly specified. The success of mitigating selection bias depends on model specification and variable measurement even if the estimation method is appropriate. See Tucker [2007, p.1079] for an example of result sensitivity to model specifications.

selection bias due to observables as well as unobservables. It is important for researchers to understand the generating process of the non-experimental data and confirm that unobservables are not the primary concern before choosing PSM.[4] Even when the primary concern is selection bias due to observables, researchers using PSM still need to identify the common support and check whether the treated and control firms matched by propensity scores are in fact close regarding the covariates. In addition, studies are advised to check the sensitivity of findings to the effects of unobservables; so far, very few do so.

The problems in IMR applications are twofold. First, some studies use the IMR method even when it is not applicable. The IMR method has been popular in accounting and finance research because of its tradition and simplicity. The method, however, is only applicable in a limited number of situations stated previously. When the IMR method is not applicable, as long as the model is parametric researchers can estimate treatment effects by full information maximum likelihood estimation (FIML). FIML is maximum likelihood estimation applied to a system of equations. It is more efficient than the IMR method even when the latter is applicable because FIML uses all information at once rather than in two steps as under the IMR approach. Second, some studies might have used wrong formulas in calculating IMR for treated and untreated (control) firms.

This article contributes to the accounting and finance literatures in several ways. First, the article discusses two popular econometric tools—PSM and IMR—in one unified framework. This structure allows researchers to compare and contrast two types of selection bias and the econometric tools that mitigate them. Second, the article provides a

[4] Even though PSM is widely used in labor economics in examining mandatory (i.e., selection by a program manager) and voluntary selections, the technique might not be as useful in accounting and finance where managers have more discretion, the decisions involve more parties, and the decision-making process is more opaque to researchers.

review of the applications of PSM at a time when the method has piqued researchers' interest. Although Roberts and Whited [2011] include a section on matching methods in their review of endogeneity in corporate finance research, their coverage is technical and does not survey PSM applications. Third, this article discusses several issues in applying the IMR method to correct selection bias due to unobservables. In contrast, the review of Francis and Lennox [2008] focuses solely on the issue of having the same covariates in both the first and second stages. Finally, compared with extant econometric readings (e.g., Heckman et al. [1998], Heckman [2001], Greene [2003], Schroeder [2010]), this article emphasizes intuition and applications and avoids unnecessary technical details, thus appealing to a broader audience in accounting and finance. Overall, the article focuses on drawing researchers' attention to the discussed issues rather than serving as a "how-to" manual. For the same reason, the survey of applications in the article is representative, but not exhaustive.

The rest of this article is organized as follows. Section 2 presents the necessary econometric framework for discussing the two types of selection bias. Section 3 introduces PSM and discusses the confusion and inadequacy in applying the method. Section 4 introduces the IMR method and discusses the confusion and mistakes in using the method. Section 5 concludes.

2. Selection Bias

This section introduces the econometric framework with binary treatment choices and defines selection bias. I use managers' decision of whether to issue a warning about a forthcoming earnings disappointment as an example. In econometric terminology, the decision to issue a warning is the "treatment" and warning firms are the "treated group;"

the decision not to issue a warning is the "non-treatment" and non-warning firms are the "untreated group." The outcome of interest is investors' price reaction to the decision to warn. Depending on whether researchers are interested in making inferences about the whole population, the subpopulation of the treated, or the subpopulation of the untreated, they may examine the "average treatment effect" (ATE), "average treatment effect on the treated" (ATT), or "average treatment effect on the untreated" (ATUT). All the three terms answer hypothetical questions. In the warning example, ATE estimates on average how stock returns differ if firms warn versus if they do not warn, *regardless of whether in reality a firm warns or not*. ATT estimates on average how the observed stock returns of *the firms that have warned* differ from the hypothetical returns had these firms not warned. ATUT estimates on average how the stock returns of *the firms that did not warn* would have been different had they warned than their observed returns without warning.

The difficulty in answering these questions is that the outcome of warning from a warning firm is of course observed, but the outcome of non-warning for the same firm is never observed and is referred to as a "counterfactual" outcome. Similarly, the outcome of a non-warning firm had it warned is also counterfactual. In non-experimental settings, researches attempt to use an observed outcome to proxy for a counterfactual outcome in estimating treatment effects. Selection bias arises when the proxy is not close to the counterfactual that is proxied for. I use the following framework to demonstrate bias in estimating ATT, since among the three treatment effects ATT draws researchers' interest most often.

Equation (1) models the outcome (Y) of treatment (with subscript 1) and Equation (2) models the outcome of non-treatment (with subscript 0), where X is the factors beyond the

treatment decision that affect the outcome and are observable to researchers and v is a collection of relevant factors unobservable to researchers.[5] Equation (3) models the treatment decision, where researchers merely observe whether a firm is treated ($T_i = 1$ when $T_i^* > 0$) or untreated ($T_i = 0$ when $T_i^* \leq 0$), not the cost-benefit analysis by the decision maker (i.e., T^* is latent). In this decision, some factors (Z) are observable to researchers and others (ε) are not.

$$Y_{1i} = \alpha_1 + X_i\beta + v_{1i} \quad \text{(Data are observed only when } T_i^* > 0\text{, that is, } T_i\text{=1)} \quad (1)$$

$$Y_{0i} = \alpha_0 + X_i\beta + v_{0i} \quad \text{(Data are observed only when } T_i^* \leq 0\text{, that is, } T_i\text{=0)} \quad (2)$$

$$T_i^* = Z_i\gamma + \varepsilon_i \quad (3)$$

By definition, ATT = E [ATT(x)], where

$$ATT(x) = \underbrace{E(Y_{1i}|_{T_i=1})}_{\text{observed}} - \underbrace{E(Y_{0i}|_{T_i=1})}_{\text{counterfactual}} \quad (4)$$

Researchers need a proxy for the counterfactual. A ready candidate for the proxy is to use an outcome observed from the untreated group (or a subset of the group). If researchers simply compare the average difference in the outcome of the treated vs. the proxied counterfactual, the estimated ATT is $\widehat{ATT} = E[\widehat{ATT}(x)]$, where

$$\widehat{ATT}(x) = E(Y_{1i}|_{T_i=1}) - \underbrace{E(Y_{0i}|_{T_i=0})}_{\text{proxy}} \quad (5)$$

The difference between true ATT and estimated ATT is the estimation bias, due to some firms being selected (or self-selected) into one group to be treated and others into the

[5] I assume that the loadings on the observables in the two outcome equations are the same, consistent with the popularly used "treatment effect model' in Greene [2003]. This assumption likely holds in general settings. For example, in the warning example, there is no reason to believe that the size and growth mimicking factors in asset pricing play different roles for warning firms than for non-warning firms. Later, I discuss studies that allow some of the loadings to vary across the outcome equations.

untreated group such that proxies have to be used for counterfactual outcomes. This bias is referred to as the "selection bias" in the econometric literature.

$$Selection\ bias = ATT - \widehat{ATT} = E[\underbrace{E(Y_{0i}\,|_{T_i=1})}_{\text{counterfactual}} - \underbrace{E(Y_{0i}\,|_{T_i=0})}_{\text{proxy}}] \tag{6}$$

It is up to a researcher to decide whether observables or unobservables are the primary cause of the selection bias, because whether a variable is observable depends on the researcher's information set. If observables, denoted by the vector of X, are the primary cause and the bias caused by unobservables can be ignored, the researcher's goal is to best control for the effects of X on the outcome either (1) by controlling and removing the effects using *a regression approach* or (2) by choosing control firms from the untreated group closest to sample firms regarding these observables using *a matching approach*. PSM is a particular matching method.

On the other hand, researchers may decide that the effects of unobservables on the estimation bias cannot be ignored. This means that the unobservables in the outcome equations, v_1 and v_0, are correlated with the unobservables in the decision (choice) model, ε (See Section 4). In other words, after researchers consider all factors that they can observe, they still believe that some unobservable factors are contributing both to the choice *and* the outcome of the treatment. For example, in the earnings warning example, perhaps managers with worse news that is unobservable to researchers but observable to some investors are more likely to warn to avoid lawsuits for withholding bad news. Although stock price plummets after the warning, leaving the impression that the act of warning has caused the price to drop, in fact the price would decline as soon as investors observe the bad news regardless of the warning. The magnitude of the price decline regardless of

warning is the selection bias due to unobservables. The IMR method is an intuitive, simple, and highly parameterized method addressing this type of selection bias. The method also conveniently controls for selection bias due to observables in the second-stage regression.

3. Propensity Score Matching and its Applications

The regression approach to mitigating selection bias due to observables imposes a linear relation between the observables ("covariates") and the outcome of interest. In addition, highly correlated covariates may induce multicollinearity in regression estimation. Matching by covariates could avoid both problems. Matched-sample designs have a long history in accounting and finance research (Cram, Karan, and Stuart [2009]; Loughran and Ritter [1997]). For example, researchers identify a control firm from the untreated group that is in the same industry as, and has the closest firm size to, the treated firm, assuming that industry and firm size are key determinants of the outcome of interest. Researchers then compare the outcomes of treated firms with those of control firms. Sometimes researchers further control for factors that are not bases of matching in a multivariate regression after initial matching.

Covariate matching is ideal when treated and untreated firms only differ on a few dimensions and the differences are represented by categorical variables. For each treated firm, researchers can find a control firm with the exact covariates. When covariate matching is feasible, it produces the best estimate of ATT (Rosenbaum and Rubin [1983]; Zhao [2004]). As the number of dimensions grows, however, covariate matching will become difficult or infeasible. That is, researchers may not find any untreated firm that shares the same characteristics as a given treated firm. PSM is one way to overcome this dimensionality problem by aggregating all covariates into one score using a likelihood

function. The use of the likelihood function may appear more sophisticated than traditional covariate matching to an audience not familiar with PSM because PSM involves estimating a parametric choice mode and calculating propensity scores. Matching by an *aggregate* score of treatment propensity, however, is driven by statistical concerns (e.g., exact covariate matching is infeasible) rather than economic concerns. This is why PSM was first proposed by statisticians and not by econometricians. The benefit of reducing dimensions from a large number of covariates to one aggregate score comes with a cost: PSM provides a coarse match (Rosenbaum and Rubin [1983]). Still, the ultimate goal of matching is to find a control group that resembles (in terms of distributional similarity) the treated group on the observable characteristics when exact matching by covariates is infeasible. Thus, it is important for researchers to check how well the treated and control firms are matched on the covariates after matching them by propensity scores. If the two groups are poorly matched on covariates, researchers may need to reconsider the specification of the likelihood function that is used in PSM in the first place.

In the framework presented in Section 2, instead of estimating the outcome regressions, researchers using PSM compare the outcomes of treated firms with those of control firms, where the control firm (or group) is identified as a firm (or subgroup) in the untreated group with a propensity score close to that of the treated firm. Here, the propensity score is the predicted likelihood of a firm being selected for treatment *often* based on the observables, X, assuming covariates $X=Z$, or on the observables that affect both the treatment and outcome (Caliendo and Kopeinig [2008], p.38). Researchers use several criteria to identify the control firm (or group) with a propensity score "close" to the treated firm's: (1) the smallest distance, (2) a group in the nearest neighborhood, or (3) kernel

weighting (where every observation in the untreated group is used with higher weights for closer observations and lower weights for more distant observations) (Diaz and Handa 2006).

The PSM estimator for ATT is often defined as the mean outcome difference of treated and control firms matched by PSM. In other words, the counterfactual outcome in Equation (4) is proxied by the average outcome of control firms selected by PSM. The estimator is unbiased under three conditions. The first condition requires that after matching by propensity scores, the selection of treatment and non-treatment can be considered random. Intuitively, it means that the selection bias is caused by observables, not unobservables (that affect both treatment selection and treatment outcome). This condition has been referred to as the "fundamental identification condition," "conditional independence" (or a weaker condition of "mean independence"), "unconfoundedness," "ignorability condition," and "selection on observables" (Zhao [2004]; Heckman et al. [1998]). The second condition requires that at the propensity scores used in matching, both treatment and non-treatment selections are possible. The condition fails at a given score if only treated firms are observable at that score. This condition is referred to as the "common support" condition. Intuitively, outside the common support, one cannot reasonably find a match for the treated firm. The third condition is balancing, that is, the distributions of covariates are approximately similar for the treated and control groups after PSM.[6]

These conditions have implications for applying PSM and drawing inferences. First, PSM is applicable to settings in which selection bias due to unobservables is *not* a major

[6] The literature of matching estimators in general lists the first two conditions. Zhao [2004, p.92] points out that these two general conditions hold under PSM only if the balancing property is satisfied. Therefore, PSM requires three conditions.

concern. There appears to be confusion among researchers about when to use PSM. The confusion is likely due to different interpretations and usage of "selection bias" and "endogeneity." "Selection bias" is technically meaningful both for selection on observables and selection on unobservables, but it is originally and more frequently used for problems of selection on unobservables (Heckman [1979]; Vella [1998]). "Endogeneity" also has different interpretations among researchers. In econometrics, "endogeneity" merely means that the covariates are correlated with the error term (Wooldridge [2002], p.50) and thus endogeneity exists in cases of selection on observables and selection on unobservables as well as in other omitted-correlated-variable situations. In other branches of economics, "endogeneity" means that a variable is determined within the context of a model—it is a choice (Wooldridge [2002], p.50). When researchers state that they use PSM to mitigate selection bias or to address endogeneity without qualifications, the statement is technically correct, but it might be interpreted by some readers as meaning that PSM *can* address selection bias due to unobservables.

Here are a few incidences of the confusion. First, the term "selection models" in econometrics refers to the econometric models dealing with selection on unobservables, not the techniques dealing with selection on observables, even though the word "selection" appears in both problems. In another example, Hamilton and Nickerson [2003] review the applications of selection models in management and suggest that selection bias is the result of treatment's being a choice. In fact, if the choice can be fully explained by variables observable to researchers, selection models are not needed even though selection bias and endogeneity technically exist. In some instances, researchers use "endogeneity" as a reason to employ PSM without arguing that the endogeneity problem is due to observables, not

unobservables (Hale and Santos [2009]; Lee and Wahal [2004]).[7] What adds to the confusion is a statement in a widely circulated literature survey by Armstrong, Guay, and Weber [2010, p.207]: "Although a propensity score research design is one technique to address identification difficulties posed by endogenous matching on unobservable variables, there are a number of other recent advances in the statistics and econometrics literatures that seem well suited to addressing this and other similar research questions." In fact, PSM and the IMR method are viewed as substitutive alternatives in some studies. For example, Doyle, Ge, and McVay [2007] report results using PSM and the IMR method side by side as alternatives to address firms' self-selection, even though their discussion apparently alludes to selection on unobservables. Lee and Wahal [2004] use an endogenous switching model, which is a variant of models addressing selection on unobservables, as an alternative to their PSM-based primary analysis. As Peel and Makepeace [2009] note, the increased popularity of PSM in recent years is perhaps due to researchers' beliefs that PSM is an alternative method to traditional Heckman procedures (for which the IMR method is the most popular implementation) for estimating treatment effects, after the latter was criticized for lacking robustness to model specifications

[7] Hale and Santos [2009, p.187] state, "In reality, this decision is likely to be endogenous. We use a propensity score matching to control for potential endogeneity of the set of firms that issues public bonds and the timing of their bond IPO." Lee and Wahal [2004, p.377] write, "Although controls are undeniably important, the crux of our problem is that venture backing is not randomly distributed, but represents an endogenous choice. This introduces a selectivity bias, one that can easily reverse inferences. To account for this bias, we use matching methods that endogenize the receipt of venture financing and do not impose linearity of function from restrictions."

(Stolzenberg and Relles [1997]).[8] However, PSM does not cure the disease that traditional Heckman procedures are supposed to cure, even though the latter is imperfect.[9]

To reduce the confusion in the literature, it is important for researchers who use PSM to be explicit that they use the technique to addresses selection on observables, not on unobservables. It is also helpful to readers if researchers could argue why selection on unobservables is not a serious concern in their setting. Appendix A lists the studies published in *Journal of Accounting and Economics* and *Journal of Financial Economics* that use PSM. It is surprising that half of the studies do not even mention that the PSM technique is for addressing selection on observables. Furthermore, when PSM is used, researchers could provide a validity check if they test the sensitivity of the findings to simulated unobservables following the procedures in Rosenbaum [2010] and Peel and Makepeace [2009]. Unfortunately, with the exception of Armstong, Jagolinzer, and Larcker [2010], most accounting and finance PSM studies do not do so.

Second, inferences from PSM are valid only for the range of propensity scores of common support. Researchers are advised to identify this range and generalize their findings to only this proportion of the population rather than to the whole population. Among accounting and finance studies that use PSM in main analyses, few check for common support and qualify the findings (Appendix A).

Third, it is important for researchers to check the balancing property, at least the means of the covariate distributions, after matching by propensity scores. The dimension

[8] Peel and Makepeace [2009] state in their opening paragraph, "While Heckman two-step model procedures are employed in this context in an endeavor to control for unobserved selection bias, researchers are becoming aware of their potential sensitivity and are increasingly turning to propensity score (PS) matching to investigate treatment effects." In fact, PSM is also sensitive to specifications of the choice model.

[9] PSM can be effectively used together with the difference-in-difference (DID) approach, where DID removes selection bias due to time-invariant unobservables so that the first condition for PSM is satisfied (Caliendo and Kopeinig [2008, p.55]; Kirk [2010]; Mclnnis and Collins [2010]; Chava and Purnanandam [2011]).

reduction by PSM is worthwhile only if the covariates of treated and control firms have similar distributions after being matched by propensity scores. If not, researchers perhaps need to modify the specification of the choice model (Caliendo and Kopeinig [2008], p.43). Among the studies surveyed in Appendix A, only 58% check the balancing property.

Other issues in applying PSM relate to the specifications of the outcome and choice models. Although a binary choice is used in the framework in Section 2, there can be more than two choices as long as they are categorical. Accordingly, propensity scores can be calculated from other discrete choice models than probit (see Armstrong et al. [2010]). Regarding the outcome equations, most studies estimate ATT merely from the average outcome difference between treated and control firms after matching; that is, the analysis is based on a univariate comparison after matching. In most accounting and finance settings, the treatment outcome may be determined by factors that do not affect treatment selection. For example, in Lee and Wahal [2004], the authors examine the differences in first-day IPO returns between venture-capital-based firms and non-VC-based firms by comparing the returns of the former with those of control firms selected from the latter group based on propensity scores of venture-capital-backing choice. General investors (other than the venture capitalists) in the IPO market probably consider other factors, such as a firm's age, business complexity, recent product development, and sales growth beyond the considerations of venture capitalists, who have provided funding at an earlier stage of the firm. Although univariate outcome comparisons after PSM produce consistent estimators of treatment effects, controlling for factors that affect treatment outcome even if they do not affect treatment selection would yield more efficient estimators.[10]

[10] Some studies include in the choice model the variables that affect the treatment outcome but not selection (e.g., underwriter ranks in Lee and Wahal [2004]). Including irrelevant variables in the choice model would

4. The Inverse Mills Ratio Method and its Applications

While PSM addresses "selection bias due to observables" by finding a control firm (T_i=0) from the untreated group that is closest to the treated firm (T_i=1) to minimize the difference between the right-hand terms of Equation (6), the IMR method addresses "selection bias due to unobservables" by estimating a bias correction term in the first stage through the choice model and adding it in the second-stage outcome regression. As the label suggests, "selection bias due to unobservables" has much to do with the unobservables in the outcome model and the unobservables in the choice model.

In Section 2 the true average treatment effect (ATE) in the parametric model is $\alpha_1 - \alpha_0$. The IMR method in fact estimates ATE, not ATT, based on observed data. Next, I will discuss how the IMR method estimates ATE and then show how researchers can infer ATT from an estimate of ATE. Still, a crude ATE estimator could be calculated by comparing the average outcome differences of treated and untreated firms, $E(Y_{1i}|_{T_i=1}) - E(Y_{0i}|_{T_i=0})$, because this is all researchers can observe. Here,

$$E(Y_{1i}|_{T_i=1}) = \alpha_1 + X_i\beta + E(\nu_{1i}|_{T_i=1}) = \alpha_1 + X_i\beta + E(\nu_{1i}|_{\varepsilon_i > -Z_i\gamma}) \tag{7}$$

$$E(Y_{0i}|_{T_i=0}) = \alpha_0 + X_i\beta + E(\nu_{0i}|_{T_i=0}) = \alpha_0 + X_i\beta + E(\nu_{0i}|_{\varepsilon_i \le -Z_i\gamma}) \tag{8}$$

Assume binormal distributions of (ν_1, ε) and (ν_0, ε) with 0 means and covariances $\sigma_{\varepsilon\nu_1}$ and $\sigma_{\varepsilon\nu 0}$ and normalize σ_ε at 1 as in binary probit models (hereinafter "probit"). Following the properties of truncated binormal distributions (Greene 2003, pp.759 and 788), we have:

$$E(\nu_{1i}|_{\varepsilon_i > -Z_i\gamma}) = \sigma_{\varepsilon\nu_1}\frac{\phi(-Z_i\gamma)}{1-\Phi(-Z_i\gamma)} = \sigma_{\varepsilon\nu_1}\frac{\phi(Z_i\gamma)}{\Phi(Z_i\gamma)}$$

---

lead to poor performance of the choice model and increase the variance of the PSM estimator (Caliendo and Kopeinig [2008], p.38).

$$E(v_{0i}\,|_{\varepsilon_i \le -Z_i\gamma}) = \sigma_{\varepsilon v_0}\frac{-\phi(-Z_i\gamma)}{\Phi(-Z_i\gamma)} = \sigma_{\varepsilon v_0}\frac{-\phi(Z_i\gamma)}{1-\Phi(Z_i\gamma)}$$

Thus, Equations (7) and (8) become (9) and (10). Differencing them yields (11):

$$E(Y_{1i}\,|_{T_i=1}) = \alpha_1 + X_i\beta + E(v_{1i}\,|_{T_i=1}) = \alpha_1 + X_i\beta + \sigma_{\varepsilon v_1}\frac{\phi(Z_i\gamma)}{\Phi(Z_i\gamma)} \tag{9}$$

$$E(Y_{0i}\,|_{T_i=0}) = \alpha_0 + X_i\beta + E(v_{0i}\,|_{T_i=0}) = \alpha_0 + X_i\beta + \sigma_{\varepsilon v_0}\frac{-\phi(Z_i\gamma)}{1-\Phi(Z_i\gamma)} \tag{10}$$

$$E(Y_{1i}\,|_{T_i=1}) - E(Y_{0i}\,|_{T_i=0}) = \underbrace{(\alpha_1 - \alpha_0)}_{\text{True ATE}} + \underbrace{[\sigma_{\varepsilon v_1}\frac{\phi(Z_i\gamma)}{\Phi(Z_i\gamma)} - \sigma_{\varepsilon v_0}\frac{-\phi(Z_i\gamma)}{1-\Phi(Z_i\gamma)}]}_{\text{Selection bias due to unobservables}} \tag{11}$$

Equation (11) indicates that our crude estimator on the left-hand side estimates ATE with bias due to unobservables. Note that the differences in observables have already been controlled for and removed by $X\beta$. To correct for the bias due to unobservables, researchers using the IMR method would estimate γ of the choice model (Equation (3)) in the first stage and add $\frac{\phi(Z_i\hat{\gamma})}{\Phi(Z_i\hat{\gamma})}$ to Equation (1) and $\frac{-\phi(Z_i\hat{\gamma})}{1-\Phi(Z_i\hat{\gamma})}$ to Equation (2), where $\hat{\gamma}$ is the estimated γ, in the second-stage least-squares regression estimations. In this way, even though treatment outcomes are observed only for a partial sample for Equation (1) and non-treatment outcomes are observed only for a partial sample for Equation (2), α_1 and α_0 can each be consistently estimated from the observed data as indicated in Equations (9) and (10). The two ratios added in the second stage are referred to as the "inverse Mills ratio" (IMR) for treated firms and untreated firms, respectively. The approach is referred to as the "two-stage least squares estimation using IMR," shortened as the "IMR method" in this article.

To implement this approach, researchers could either estimate the augmented Equations (1) and (2) separately or stack them into one equation, Equation (12), using an indicator variable T to distinguish treatment from non-treatment ("treatment" takes the value of 1) and setting the dependent variable to be $Y_i = Y_{1i} * T_i + Y_{0i} * (1-T_i)$. The coefficient on T is then the estimated ATE and the coefficients on the IMR variables are the estimated covariance between the unobservables in the treatment decision and those in the treatment outcome regression. Equation (12) is presented in Tucker [2007] and is modified from the standard treatment-effect model of Greene [2003, p.788], which constrains the coefficients on the IMR for treated and untreated firms (i.e., $\sigma_{\varepsilon v1}$ and $\sigma_{\varepsilon v_0}$) to be the same. Wooldridge [2002, p.631] instead presents a more flexible model than Equation (12) by allowing the coefficients on the exogenous observables to differ across the treatment outcome and non-treatment outcome equations.

$$Y_i = \alpha_0 + \underbrace{(\alpha_1 - \alpha_0)}_{\text{ATE}} * T_i + X_i\beta + \sigma_{\varepsilon v_1} IMR_i * T_i + \sigma_{\varepsilon v_0} IMR_i * (1 - T_i) + w_i \qquad (12)$$

With the estimate of ATE, ATT can be estimated according to the relation:[11]

$$ATT = ATE + E[E(Y_{1i} \mid_{T_i=1}) - E(Y_{0i} \mid_{T_i=1})] = ATE + (\sigma_{\varepsilon v_1} - \sigma_{\varepsilon v_0}) * E[\frac{\phi(Z_i\gamma)}{\Phi(Z_i\gamma)}]$$

It might appear odd that biases caused by *unobservables* can ever be estimated and controlled. The intuition behind it is as follows. In estimating average treatment effects,

---

11

$$ATT(x) = E(Y_{1i} \mid_{T_i=1}) - E(Y_{0i} \mid_{T_i=1}) = [\alpha_1 + X_i\beta + E(v_{1i} \mid_{T_i=1})] - [\alpha_0 + X_i\beta + E(v_{0i} \mid_{T_i=1})]$$

$$= (\alpha_1 - \alpha_0) + E(v_{1i} \mid_{T_i=1}) - E(v_{0i} \mid_{T_i=1}) = (\alpha_1 - \alpha_0) + \sigma_{\varepsilon v_1} \frac{\phi(Z_i\gamma)}{\Phi(Z_i\gamma)} - \sigma_{\varepsilon v_0} \frac{\phi(Z_i\gamma)}{\Phi(Z_i\gamma)}$$

$$= ATE + (\sigma_{\varepsilon v_1} - \sigma_{\varepsilon v_0}) \frac{\phi(Z_i\gamma)}{\Phi(Z_i\gamma)}$$

See Wooldridge [2002, pp. 606-.607] and Schroeder [2010, p.215].

even though researchers do not observe ε, v_1, and v_0, all researchers need to know is the *mean* effect of the unobservable factors in the treatment decision on the treatment outcome given observed data. This effect can be calculated from truncated bivariate distributions of the unobservables as long as the distributions are specified. The latter condition is automatically satisfied in parametric analysis. For example, in ordinary least squares, even though the error term is not observable, we make specific assumptions about its distribution. Thus, this mean effect is estimated from the first stage and added to the second stage for error correction. This is the intuition of all two-stage estimations of selection models, where the IMR method is a special, albeit restrictive, case.

As Equation (11) shows, the selection bias to be corrected by the IMR method has two components: one related to the treated group and the other related to the untreated group. The magnitude of each component increases with the covariance between the unobservables in the choice model and the unobservables affecting treatment outcome. Of course, when these unobservables are not correlated, there is no selection bias from unobservables. These covariances are estimated from the second-stage regression. It might be surprising that the IMRs contribute to the bias even though they are made of observables in the choice model. The contribution is because the IMRs reveal information about the unobservables of the treatment decision. In particular, given the observed choice, one can infer about the unobservable based on the observables because the two together in Equation (3) determine whether the net benefits cross a threshold for managers to select the treatment.[12]

[12] The IMRs are monotonic transformations of $Z_i\gamma$ with reversing ordering (See Tucker [2007]).

Treatment-effect models are by far the most common selection models in accounting and finance research. Researchers occasionally use other selection models, for example, the "traditional Heckman" model (Heckman [1979]) when they are interested in the *determinants* of the outcome after treatment or non-treatment but not the difference in outcome between treatment and non-treatment. When the treatment decision depends on the perceived outcome of treatment vs. non-treatment, researchers use the "endogenous switching" model. The IMR method could be used for all the three models.

The IMR method is simple but highly parameterized. Its simplicity perhaps explains why it is widely used in accounting and finance research.[13] However, the method requires strong assumptions for both the outcome regression and the choice model. The error correction variable is in the form of IMR only when (1) the outcome regression is linear, (2) the choice model is probit, and (3) the unobservables in the first and second stages follow bivariate normal distributions. When these requirements are not satisfied, the error correction variable will be in other forms and adding IMR to the outcome regression will not correct the selection bias that researchers intend to correct.

In Figure 1 I summarize various ways in which researchers use selection models. The IMR method is applicable to only the first three settings: treatment-effect model, traditional Heckman, and endogenous switching. In these settings the IMR estimators are consistent, but their standard errors must be adjusted for sampling errors that occur in the first stage (Greene [1981]; Maddala [1983]). The adjustments are nontrivial and are best done using statistical software. Moreover, the IMR method is less efficient than the full information maximum likelihood estimation (FIML), where the latter uses maximum

[13] A keyword search of "inverse mills ratio" yields 24 articles in *The Accounting Review*, 12 in *Journal of Accounting Research*, 16 in *Journal of Accounting and Economics*, 9 in *Journal of Finance*, 36 in *Journal of Financial Economics*, and 9 articles in *Review of Financial Studies*.

likelihood estimation on a system of equations.[14] In settings for which the IMR method is not appropriate, FIML is applicable as long as the models are parametric.[15] I discuss examples of accounting applications below to highlight these points.

Some studies use the IMR method even though the second-stage outcome model is discrete and therefore nonlinear. For example, Weber and Willenborg [2003] examine in probit in the second stage whether a firm's pre-IPO auditor's opinion is more predictive of its post-IPO stock survival when the auditor is a Big-6 firm than when it is not and model the Big-6 decision in the first stage. Wu and Zang [2009] model financial analysts' departures vs. stays after brokerage mergers in the first stage and examine in two separate logit models in the second stage the internal promotion of analysts to research executive positions after they stay and the external promotion of analysts after they leave the original brokerage firm. Cohen and Zarowin [2010] examine in probit in the second stage the likelihood of real earnings management at firms that have already been identified in the first stage as having managed earnings through either real or accrual management**.** All the three studies estimate a probit model in the first stage and add IMR to the second stage, intending to correct for selection bias due to unobservables However, the IMR term(s) does not correct for the bias because the second-stage model is nonlinear. A bivariate probit model with sample selection would be appropriate (Greene [2002], p.E17-19). The model can be estimated by FIML.[16]

---

[14] Among the accounting studies surveyed, Omer, Bedard, and Falsetta [2006] and Li and Ramesh [2009] use FIML. The latter acknowledges the efficiency of FIML over the IMR method.

[15] FIML does not need additional assumptions beyond those for two-stage estimations and is not necessarily less robust to violations of assumptions than are two-stage estimations. Two-stage estimations, on the other hand, require less computing power than FIML. Perhaps for this reason, two-stage estimations were more popularly used than FIML in the early years.

[16] Terza [2009, p.563] demonstrates that two-stage estimations are feasible even when the second-stage model is nonlinear. The error correction variables to be added in the second-stage are in complicated forms, not in the form of IMR. Nonlinear least squares estimation is used for the augmented second-stage equation.

Some studies use the IMR method even though the choice model is logit, not probit. For example, Feng and Koch [2010], Engle, Hayes, and Wang [2007], and Khurana and Raman [2004] use logit in the first stage and add IMR to the second-stage linear regression. This is inappropriate because IMR requires the arguments in the p.d.f. in the numerator and c.d.f. in the denominator of the ratio to be normally distributed. When logit is used in the first stage, one cannot simply use the IMR formulas, but needs to transform $Z_i\gamma$ by an inverse standard normal c.d.f. function to ensure that the arguments in the numerator and denominator functions are normally distributed, not logistically distributed (Greene [2002], p.E23-71).

In some studies there are in fact more than two choices in the first stage; therefore, a probit model is insufficient. For example, Rogers [2008] examines whether disclosure quality differs if managers subsequently trade stocks than if managers' trading incentives are absent and, more importantly, whether disclosure quality is higher before insider selling than before insider buying. Insiders have three choices: "sell," "hold," and "buy" the firm's stocks. Perhaps for econometric convenience, in the first stage the author assumes that managers consider either "sell vs. hold" or "buy vs. hold" and models "sell" vs. "hold" and "buy" vs. "hold" separately in probit. This research design essentially changes managers' decision of three choices to two sequential decisions, the first of which (i.e., deciding to go down the path of "sell vs. hold" or the other path of "buy vs. hold") is skipped and never modeled in Rogers [2008].

More than two choices of treatment are common in accounting and finance research. In these situations, researchers often use ordered probit for ordered choices and multinomial logit for unordered choices in the first stage and model treatment outcome in a linear

regression in the second stage. When the second-stage model is linear, researchers can estimate the choice model in the first stage, calculate the expected outcome residual given each choice, and add the variables to the second stage as long as the outcome of the treatment choice is observed. These new variables are ratios but are not IMRs (see Vella [1998], pp.147-148; Greene [2002], p.E23-79). In addition, an inverse standard normal c.d.f. transformation is required when multinomial logit is used in the first stage (see Greene [2002], pp. E23-72 to 73).

There are four other issues in applying the IMR method. First, the formula to calculate IMR for treated firms is different from that for untreated firms, even though one variable label is used for both groups in standard treatment-effect models. The IMRs for treated and untreated firms have opposite signs. This is why the IMR variable is highly correlated with the treatment indicator variable. This high correlation is not a weakness of the selection model. Second, there have been two sets of formulas for IMR in the econometric literature and the difference between them is the sign. The IMR formulas described above can be found in Heckman [1979], Vella [1998], and Greene [2003]. But in some other studies the formulas are $-\frac{\phi(Z_i\gamma)}{\Phi(Z_i\gamma)}$ for treated firms and $\frac{\phi(Z_i\gamma)}{1-\Phi(Z_i\gamma)}$ for untreated firms (Lee [1978]; Maddala [1983], p.223-224). The differences are due to the sign of the error term in the choice model: it is positive in the former literature but negative in the latter literature (i.e., $T_i^* = Z_i\gamma - \varepsilon_i$). Because popular software (SAS, Stata, and Limdep) uses a positive sign for the error term, the formulas in this article are consistent with model estimations from modern software. If researchers are merely interested in the average treatment effect, the sign differences in IMR calculations will not affect the estimate of the effect. But if researchers also make inferences about the correlation of the unobservables in the

treatment decision and treatment outcome, a wrong sign for IMR will result in a wrong sign for the estimated correlation.

This detail of IMR formulas is sometimes missed by researchers. For example, Hamilton and Nickerson [2003] survey and propose sophisticated selection model techniques for the management literature, but the formulas they provide in their text and Appendix 1 are incorrect. They specify a positive sign for the error term in the first stage but follow the formulas in the latter econometric literature that uses the error term with a negative sign. Early accounting studies that use the IMR method provide correct formulas (Core and Guay [1999]; Verrecchia and Leuz [2000]). Most subsequent studies, with the exceptions of Chaney, Jeter, and Shivakumar [2004] and Tucker [2007], are not explicit about the formulas for IMR calculations and thus there is no telling whether the formulas used for treated and untreated firms are different and correct. From the glimpse of discussion provided, confusion about the IMR formulas seems to continue. For example, Chen, Matsumoto, and Rajgopal [2010] note that their IMR for treated firms is a *negative* function of fitted probabilities from the choice model, implying that their formula is incorrect because they use a positive sign for the error term in the first stage (they do make inference about the correlation of unobservables). Given inconsistent notations in the econometric literature and confusion in applications, future research could be explicit and cautious about IMR calculations.[17]

[17] Haw, Hu, Hwang, and Wu [2004, p.439] note, "Including the expected probability of being in the final sample (*inverse Mills ratio*) as another explanatory variable in the regressions does not alter the findings." This statement implies misunderstanding of IMR. Hogan [1997] uses an endogenous switching model to examine a firm's tradeoff between the benefits of less IPO underpricing and the cost of higher audit fees from hiring Big-6 auditors. Her hypotheses rely upon the signs of the estimated correlations of the unobservables, but the signs of her IMRs are incorrect, because the formula for IMR in her Equation (1) is based on choice model $T_i^* = Z_i\gamma - \varepsilon_i$, but her empirical choice model, her Equation (4), is $T_i^* = Z_i\gamma + \varepsilon_i$.

The third issue is other variations of the treatment-effect model. One variation is to allow the coefficients on the exogenous covariates, *X*, to differ for treated firms and untreated firms. This can be achieved by adding interaction terms of the treatment indicator and the exogenous covariates. Examples of such models are Givoly, Hayn, and Katz [2010], Louis [2005], Leone, Rock, and Willenborg [2007], and Chung and Wynn [2008]. Another variation is to examine two treatments (not two choices of one treatment). When the treatments are independent, the second-stage linear regression can include both indicator variables for the treatments and both IMRs, each calculated separately from a probit choice model for the treatment. When the treatments are dependent, the first stage requires a biprobit model, estimated simultaneously. Two ratios will then be calculated involving marginal binormal p.d.f. in the numerator and binormal c.d.f. in the denominator (Greene [2002], pp.E23-83, E17-15), not in the form of IMRs. The ratios are then added to the second stage, corresponding to the two treatment indicator variables capturing average treatment effects, for the two-stage least squares estimation. Of course, the model can be estimated by FIML.

Examples of two-treatment-effect studies are Muller and Riedl [2002] and Asquith, Beatty, and Weber [2005]. Muller and Riedl [2002] examine the choice of Big 6 vs. non-Big 6 and the choice of external property appraiser vs. internal appraiser together in the first stage. They then add two IMR ratios to the second stage. Asquith, Beatty, and Weber [2005] examine the choices of interest-decreasing performance pricing and interest-increasing performance pricing bank debt contracts in the first stage and interest rate spreads in the second stage. They estimate the two stages simultaneously in the primary

analysis and use the two-stage least squares IMR method for robustness. Neither study, however, indicates that the ratios they use are not traditional IMRs.

The last issue is having the same covariates in the choice model and the treatment outcome regression. Francis and Lennox [2008] discuss this issue in depth. Among the accounting studies I survey, one study uses the same covariates in both stages. Core and Guay [1999] examine firms' decisions of granting options in the first stage and examine in the second stage whether the size of the option grant is negatively associated with the firm's deviation of equity incentives from an optimal level. Economy theories should determine what covariates belong to each stage. When the treatment choice and treatment outcome are distinctive economic decisions, the covariates in the two stages are probably different. If the covariates in the two stages are the same, the identification in the second stage can be weak, the two-stage least squares approach can be unreliable, and the findings should be interpreted with caution (Vella [1998], p.135; Wooldridge, [2002], p.564).

5. Conclusion

This article discusses two popular econometric techniques that are receiving growing interest in accounting and finance research—the propensity-score matching method and the two-stage least-squares inverse-Mills-ratio method. The former addresses selection bias due to observables and the latter addresses selection bias due to unobservables. I discuss the conditions under which each method can be properly used as well as the confusion, inadequacy, and perhaps mistakes in their applications. The discussion assumes that researchers have already identified and properly measured all observable factors and that the form of their relations is correctly specified. One can never overemphasize the importance of identifying and properly measuring observable factors, because the more

successfully a researcher carries out this task, the less challenge he/she faces from the thornier problem of selection on unobservables (than the problem of selection on observables).

The article has four takeaways. First, researchers should use the proper tool for a given problem: PSM mitigates selection bias due to observables, but does not address selection bias due to unobservables. Second, in applying PSM researchers are advised to test the differences in distributions of covariates between treated and control firms matched by propensity scores and to restrict their inferences to firms whose characteristics can be found in both the treated and control groups. The advantage of PSM over other matching methods is dimension reduction. Matching by PSM, however, does not necessarily guarantee that treated and control firms are well matched by covariates. Inferences are invalid outside the range where good matches cannot be found. Third, the IMR method, which is a simple, popular, and restrictive case of two-stage least squares estimation, is limited to situations in which the outcome of interest is modeled in a linear regression, the choices are binary, and the error terms follow bivariate normal distributions. Many studies use the IMR method despite violations of these conditions. As long as the models are parametric, researchers can overcome these problems either by generalized two-stage least-squares or nonlinear least-squares estimations with error correction variables in the form of ratios *other than* IMR to be added in the second stage or by full information maximum likelihood estimation (FIML) The latter is more efficient than the former because FIML uses all information at once rather than in two steps. Even when the IMR method is applicable, FIML yields more efficient estimators than the IMR method for the same reason why FIML is more efficient than two-stage estimations. Last, when the IMR

method is used, researchers are advised to be explicit about the formulas because of the confusion in the literature on this point.

The methods discussed in the article are within the realm of frequentists' parametric framework. This framework is by far the most commonly used by accounting and finance researchers. On the other hand, nonparametric methods—which allow the covariates to have unknown functional form and the error terms to have unknown distributions—have been proposed to evaluate treatment effects when selections are on observables (Imbens [2004]) or when selections are on unobservables (Das, Newey, and Vella [2003]; Heckman and Vytlacil [2005]). In recent years Bayesian analysis has been revolutionized by the increased computing power of computers and the development of Markov chain Monte Carlo (McMC) stochastic integration methodology (Carlin and Chib [1995]). Econometricians have proposed Bayesian methods to evaluate treatment effects when selections are on unobservables (Li, Poirier, and Tobias [2004]; Chib [2007]; Chib and Jacobi [2007]).[18] To my knowledge, so far no accounting and finance archival study has used nonparametric or Bayesian methods to examine the treatment effects of managers' decisions. As Greene [2003, p.708] points out, "the fewer assumptions one makes about the population, the less precise the information that can be deducted by statistical techniques." At this point little is known (at least to me) about the gains of nonparametric and Bayesian methods over traditional parametric methods. Future research may explore, employ, and evaluate statistical methods in these frameworks and examine when alternative methods make a difference and provide *new* insights into the consequences of managers' corporate decisions.

[18] See Schroeder [2010] for extensive coverage of the Bayesian perspective as well as simulated examples.

REFERENCES

Armstrong, S. C., W. R. Guay, and J. P. Weber. 2010. The role of information and financial reporting in corporate governance and debt contracting. *Journal of Accounting and Economics* 50 (2-3): 179-234.

Armstrong, S. C., A. D. Jagolinzer, and D. F. Larcker. 2010. Chief executive officer equity incentives and accounting irregularities. *Journal of Accounting Research* 48 (2): 225-271.

Asquith, P., A. Beatty, and J. Weber. 2005. Performance pricing in bank debt contracts. *Journal of Accounting and Economics* 40 (1-3); 101-128.

Bae, K-H., J-K Kang, and J. Wang. 2010. Employee treatment and firm leverage: A test of the stakeholder theory of capital structure. *Journal of Financial Economics.* Forthcoming.

Blouin, J., J. E. Core, and W. Guay. 2010. Have the tax benefits of debt been overestimated? *Journal of Financial Economics* 98 (2):195-213.

Bottazzi, L., M. D. Rin, and T. Hellmann. 2008. Who are the active investors? Evidence from venture capital. *Journal of Financial Economics* 89 (3) 488-512.

Caliendo, M. and S. Kopeing. 2008. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Survey* 22 (1): 31-72.

Carlin, B. P. and S. Chib. 1995. Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society* 57 (3): 473-484.

Campello, M., J. R. Graham, and C. R. Harvey. 2010. The real effects of financial constraints: Evidence from a financial crisis. *Journal of Financial Economics* 97 (3): 470-487.

Chaney, P. K., D. Jeter, and L. Shivakumar. 2004. Self-selection of auditors and audit pricing in private firms. *The Accounting Review* 79 (1): 51-72.

Chava, S. and A. Purnanandam. 2011. The effect of banking crisis on bank-dependent borrowers. *Journal of Financial Economics* 99 (1): 116-135.

Chib, S. 2007. Analysis of treatment response data without the joint distribution of potential outcomes. *Journal of Econometrics* 140: 401-412.

Chib, S. and L. Jacobi. 2007. Modeling and calculating the effect of treatment at baseline from panel outcomes. *Journal of Econometrics* 140: 781-801.

Chen, S., D. Matsumoto, and S. Rajgopal. 2010. Is silence golden? An empirical analysis of firms that stop giving quarterly earnings guidance. *Journal of Accounting and Economics.* Forthcoming.

Chung, H. H. and J. P. Wynn. 2008. Managerial legal liability coverage and earnings conservatism. *Journal of Accounting and Economics* 46 (1); 135-153.

Cohen, D. A. and P. Zarowin. 2010. Accrual-based and real earnings management activities around seasoned equity offerings. *Journal of Accounting and Economics* 50: 2-19.

Core, J. and W. Guay. 1999. The use of equity grants to manage optimal equity incentive levels. *Journal of Accounting and Economics* 28 (2); 151-184.

Cram, D. P., V. Karan, and I. Stuart. 2009. Three threats to validity of choice-based and matched-sample studies in accounting research. *Contemporary Accounting Research* 26 (2): 477-516.

Das, M., W. K. Newey, and F. Vella. 2003. Nonparametric estimation of sample selection models. *The Review of Economic Studies* 70 (1): 33-58.

Diaz, J. J. and S. Handa. 2006. An assessment of propensity score matching as a nonexperimental impact estimator: Evidence from Mexico's PROGRESA program. *Journal of Human Resources* 4 (2): 319-345.

Doyle, J. T., W. Ge, and S. McVay. 2007. Accruals quality and internal control over financial reporting. *The Accounting Review* 82 (5); 1141-1170.

Engle, E., and R. M. Hayes, and X Wang. 2007. The Sarbanes-Oxley Act and firms' going-private decisions. *Journal of Accounting and Economics* 44 (1-2); 116-145.

Faulkender, M. and J. Yang. 2010. Inside the black box: The role and competition of compensation peer groups. *Journal of Financial Economics* 96 (2): 257-270.

Feng, M. and A. S. Koch. 2010. Once bitten, twice shy: The relation between outcomes of earnings guidance and management guidance strategy. *The Accounting Review* 85 (6): 951-1984.

Givoly, D., C. K. Hayn, and S. P. Katz. 2010. Does public ownership of equity improve earnings quality? *The Accounting Review* 85 (1): 195-225.

Greene, W. H. 1981. Sample selection bias as a specification error: Comment. *Econometrica* 49 (3); 795-798.

Greene, W. H. 2002. *Limdep Manual*. Version 8.0. Econometric Software, Inc.

Greene, W. H. 2003. *Econometric Analysis*. 5th edition. Prentice Hall.

Francis, J. R. and C. S. Lennox. 2008. Selection models in accounting research. Hong Kong University of Science and Technology.

Hale, G. and J. A. C. Santos. 2009. Do banks price their informational monopoly? *Journal of Financial Economics* 93 (2): 185-206.

Hamilton, B. H. and J. A. Nickerson. 2003. Correcting for endogeneity in strategic management research. *Strategic Organization* 1 (1): 51-78.

Haw, I-M., B. Hu, L-S. Hwang, and W. Wu. 2004. Ultimate ownership, income management, and legal and extra-legal institutions. *Journal of Accounting Research* 42 (2); 423-462.

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica* 47 (1): 153-161.

Heckman, J. J., H. Ichimura, and P. Todd. 1998. Matching as an econometric evaluation estimator. *Review of Economic Studies* 65 (2): 261-294.

Heckman, J. J. 2001. Microdata, heterogeneity and the evaluation of public policy. *Journal of Political Economy* 109 (4): 673-748.

Heckman, J. J. and E. Vytlacil. 2005. Structural equations, treatment effects, and econometric policy evaluations. *Econometrica* 73 (3): 669-738.

Hillion, P. and T. Vermaelen. 2004. Death spiral convertibles. *Journal of Financial Economics* 71 (2): 381-415.

Hogan, C. E. 1997. Costs and benefits of audit quality in the IPO market: A self-selection analysis. *The Accounting Review* 72 (1); 67-86.

Imbens, G. W. 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics* 86 (1): 4-29.

Kirk, M. 2010. Research for sale: Determinants and consequences of paid-for analyst research. *Journal of Financial Economics.* Forthcoming.

Khurana, I. K. and K. K. Raman. 2004. Litigation risk and the financial reporting credibility of Big 4 versus non-Big 4 audits: Evidence from Anglo-American countries. *The Accounting Review* 79 (2): 473-495.

Lee, P. M. and S. Wahal. 2004. Grandstanding, certification and the underpricing of venture capital backed IPOs. *Journal of Financial Economics* 73 (2): 375-407.

Lee, L. 1978. Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables. *International Economic Review* 19 (2): 415-433.

Leone, A. J., S. Rock, and M. Willenborg. 2007. Disclosure of intended use of proceeds and underpricing of initial public offerings. *Journal of Accounting Research* 45 (1): 111-153.

Li, M., D. J. Poirier, and J. L. Tobias. 2004. Do dropouts suffer from dropping out? Estimation and prediction of outcome gains in generalized selection models. *Journal of Applied Econometrics* 19 (2): 203-225.

Li, E. and K. Ramesh. 2009. Market reaction surrounding the filing of periodic SEC reports. *The Accounting Review* 84 (4): 1171-1208.

Loughran, T. and J. R. Ritter. 1997. The operating performance of firms conducting seasoned equity offerings. *Journal of Finance* 52 (5): 1823-1850.

Maddala, G. S. 1983. *Limited-dependent and qualitative variables in econometrics.* Cambridge.

Massoud, N., D. Nandy, A. Saunders, and K. Song. 2010. Do hedge funds trade on private information? Evidence from syndicated lending and short-selling. *Journal of Financial Economics.* Forthcoming.

Mclnnis, J. and D. W. Collins. 2010. The effect of cash flow forecasts on accrual quality and benchmark beating. *Journal of Accounting and Economics*. Forthcoming.

Muller, K. A. and E. J. Riedl. 2002. External monitoring of property appraisal estimates and information asymmetry. *Journal of Accounting Research* 40 (3): 865-881.

Murphy, K. J. and T. Sandino. 2010. Executive pay and "independent' compensation consultants. *Journal of Accounting and Economics* 49 (3): 247-262.

Omer, T. C., J. C. Bedard, and D. Falsetta. 2006. Auditor-provided tax services: The effects of a changing regulatory environment. *The Accounting Review* 81 (5): 1095-1117.

Officer, M. S., O. Ozbas, and B. A. Sensory. 2010. Club deals in leveraged buyouts. *Journal of Financial Economics* 98 (2): 214-240.

Ovtchinnikov, A. V. 2010. Capital structure decisions: Evidence from deregulated industries. *Journal of Financial Economics* 95 (2): 249-274.

Peel, M. J. and G. H. Makepeace. 2009. Propensity score matching in accounting research and Rosenbaum bounds analysis for confounding variables. Working paper. Cardiff University, U.K. SSRN.

Rosenbaum, P. R. 2010. *Observational Studies*. Springer. 2nd edition.

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

Rogers, J. L. 2008. Disclosure quality and management trading incentives. *Journal of Accounting Research* 46 (5): 1265-1296.

Schroeder, D. A. 2010. *Accounting and causal effects: Econometric challenges*. Springer.

Stolzenberg, R. M. and D. A. Relles. 1997. Tools for intuition about sample selection bias and its correction. *American Sociological Review* 62 (3); 494-507.

Stuart, T. E. and S. Yim. 2010. Board interlocks and the propensity to be targeted in private equity. *Journal of Financial Economics* 97 (1): 174-189.

Terza, J. V. 2009. Parametric nonlinear regression with endogenous switching. *Econometric Review* 28 (6): 555-580.

Tucker, J. W. 2007. Is openness penalized? Stock returns around earnings warnings. *The Accounting Review* 82 (4): 1055-1087.

Vella, F. 1998. Estimating models with sample selection bias: A survey. *The Journal of Human Resources* 33 (1): 127-169.

Verrecchia, R. E. and C. Leuz. 2000. The economic consequences of increased disclosure. *Journal of Accounting Research* 38 (3): 91-124.

Villalonga, B. and R. Amit. 2006. How do family ownership, control and management affect firm value? *Journal of Financial Economics* 80 (2): 385-417.

Weber, J. and M. Willenborg. 2003. Do expert informational intermediaries add value? Evidence from auditors in microcap IPOs. *Journal of Accounting Research* 41 (4): 681-720.

Wooldridge, J. M. 2002. *Econometric analysis of cross section and panel data*. The MIT Press.

Wu, J. S. and A. Y. Zang. 2009. What determine financial analysts' career outcomes during mergers? *Journal of Accounting and Economics* 47 (1-2): 59-86.

Zhao, Z. 2004. Using matching to estimate treatment effects: data requirement, matching metrics, and Monte Carlo evidence. *Review of Economics and Statistics* 86 (1): 91-107.

APPENDIX A: PSM Applications in JAE and JFE

No.	Paper	Purpose	A	B	C	D	E
1	Mclnnis and Collins (2010)	Examine the effect of analysts' issuing cash flow forecasts on firms' earnings management and meeting analyst earnings expectations in a DID design	Primary	Y	-	N	Y
2	Murphy and Sandino (2010)	Examine the effect of management's directly hiring compensation consultants on executive pay	Primary	Y	N	N	Y
3	Hillion and Vermaelen (2004)	Examine whether firms use floating-priced convertibles as a last-resort financing tool by comparing performance changes of issuing firms with non-issuing firms	Primary	Y	N	N	N
4	Lee and Wahal (2004)	Examine the effect of venture capital backing on the first-day returns of IPOs	Primary	N	N	N	N
5	Villalonga and Amit (2006)	Examine the effect of founder-CEO or descendent-CEO on equity value of family firms (compared with that of nonfamily firms)	Robustness	Y	-	-	-
6	Bottazzi et al. (2008)	Examine the effect of investor activism on firm performance using the IMR method	Robustness	Y	-	-	-
7	Hale and Santos (2009)	Examine the effect of bond IPO on the interest spread of bank loans	Primary	N	N	N	N
8	Bae et al. (2010)	Examine the effect of high employee-friendly ratings on a firm's leverage ratio	Robustness	N	N	Y	Y
9	Blouin et al. (2010)	Use PSM to deal with survival issues: filling future-year income of missing firms by that of non-missing firms matched by the probability of survival	Robustness	N	-	-	-
10	Campello et al. (2010)	Examine the effect of financial constraints on corporate spending during the financial crisis	Primary	Y	N	N	Y
11	Faulkender and Yang (2010)	Examine whether firms select highly paid peers to justify their own CEO pay	Primary	N	N	N	N
12	Kirk (2010)	Examine the effect of paid-for analyst research on a firm's information environment	Primary	N	-	N	Y
13	Massoud et al. (2010)	Examine the effect of loan origination by hedge funds on short-selling of firm stock	Robustness	Y	N	N	Y
14	Ovtchinnikov (2010)	Examine the effect of deregulation on firm leverage	Robustness	N	-	-	-
15	Officer et al. (2010)	Examine the effect of club deal LBO on the price premium of buyouts	Robustness	N	N	N	N
16	Stuart and Yim (2010)	Examine the effect of board interlocks on the likelihood of being targeted in private equity transactions	Robustness	N	N	-	-
17	Chava and Purnanandam (2011)	Examine the effect of bank dependence on firm value during market-wide negative capital shocks in a DID design	Primary	Y	-	N	Y

Note: "DID" means that "difference-in-difference." "-" is used for all the last three columns when PSM is discussed briefly without results being presented. "-" is also used when a question is not applicable because of the specific research design indicated below.
A: Is PSM used in primary analysis or robustness tests?
B: Does the paper make it clear that PSM is applicable to "selection on observables"? This question is not applicable when DID design is used.
C: Does the paper examine the sensitivity of results to "selection on unobservables"?
D: Does the paper identify common support? This question is not applicable when kernel weighting is used in matching.
E: Does the paper examine the effectiveness of propensity score matching by testing the difference in the means of sample and control firms' covariates?

FIGURE 1: Settings of Selection Bias due to Unobservables

Settings	1st stage	2nd stage	2nd stage obs. in relation to 1st stage obs.	Two-stage estimation	Superior estimation
Treatment effect	Probit	Linear reg.	Same	Add IMR to 2nd stage	FIML
Traditional Heckman	Probit	Linear reg.	Subset	Add IMR to 2nd stage	FIML
Endogenous switching	Probit	Linear reg.	Subset	Add IMR to 2nd stage	FIML
Variations:					
A	logit	Linear reg.	Subset or same	Use inverse normal c.d.f. transformation and add ratios to 2nd stage	FIML
B	tobit	Linear reg.	Subset	Add tobit residuals to 2nd stage	FIML
C	Ordered probit	Linear reg.	Subset or same	Add ratios other than IMR to 2nd stage	FIML
D	Multinomial logit	Linear reg.	Subset or same	Add ratios other than IMR to 2nd stage	FIML
E	biprobit	Linear reg.	Subset or same	Add ratios other than IMR to 2nd stage	FIML
F	Probit	Ordered probit	Subset	Feasible but not advisable	FIML
G. Biprobit with selection	probit	probit	Subset	Feasible but not advisable	FIML
H	probit	tobit	Subset	Feasible but not advisable	FIML

Note: With "full information maximum likelihood estimation" (FIML), the equations are estimated together in one system. The full version of Limdep handles all the above settings and many more.