

Selection Models in Accounting Research

Clive S. Lennox

Nanyang Technological University

Jere R. Francis

University of Missouri–Columbia

Zitian Wang

Nanyang Technological University

ABSTRACT: This study explains the challenges associated with the Heckman (1979) procedure to control for selection bias, assesses the quality of its application in accounting research, and offers guidance for better implementation of selection models. A survey of 75 recent accounting articles in leading journals reveals that many researchers implement the technique in a mechanical way with relatively little appreciation of important econometric issues and problems surrounding its use. Using empirical examples motivated by prior research, we illustrate that selection models are fragile and can yield quite literally any possible outcome in response to fairly minor changes in model specification. We conclude with guidance on how researchers can better implement selection models that will provide more convincing evidence on potential selection bias, including the need to justify model specifications and careful sensitivity analyses with respect to robustness and multicollinearity.

Keywords: *selection model; Heckman; selection bias; endogeneity; treatment effect model.*

Data Availability: *Data used are available from public sources identified in the study.*

I. INTRODUCTION

This study evaluates the implementation of selection models in the accounting literature, provides guidance to accounting researchers on potential problems with selection models, and recommends some steps that can be taken to improve their implementation. Such guidance is especially important given the increased use of selection models and the frequent comments by editors and reviewers of the need to control for endogeneity and selection bias. Over

We thank Steven Kachelmeier (editor) and the two reviewers for their helpful comments throughout the review process. We also appreciate the comments of Mark Clatworthy, Bill Griffiths, Gilles Hilary, David Larcker, Christian Leuz, Siu Fai Leung, Ping-Sheng Koh, David Maber, Chul Park, Mike Peel, Jeff Pittman, and Terry Shevlin, and the research assistance of Rui Ge and Scott Seavey.

Editor's note: Accepted by Steven Kachelmeier.

Submitted: April 2010

Accepted: July 2011

Published Online: November 2011

the period 2000 through 2009, we identify 75 articles from *The Accounting Review*, *Journal of Accounting and Economics*, *Journal of Accounting Research*, *Contemporary Accounting Research*, and *Review of Accounting Studies* that use selection models out of 1,016 empirical articles published in these journals over the same period. The recent trend is even stronger with 11 percent of empirical articles employing a selection model during 2008 to 2009.

Selection occurs when observations are non-randomly sorted into discrete groups, resulting in the potential for coefficient bias in estimation procedures such as ordinary least squares (OLS) (Maddala 1991). The standard approach to controlling for selection bias is the procedure developed by Heckman (1979), hereafter referred to as the selection model. A convincing implementation requires the researcher to identify exogenous independent variables from the first stage choice model that can be validly excluded from the set of independent variables in the second stage regression (Little 1985). However, the importance of exclusion restrictions appears to have fallen under the radar of the accounting literature. A surprising number of studies (14 of 75) fail to have any exclusions, and other studies (7 out of 75) do not report the first stage model, making it impossible to determine if they imposed exclusion restrictions. Moreover, very few studies provide any theoretical or economic justification for their chosen restrictions.

We demonstrate empirically that the selection model is fragile and that results can be non-robust and therefore unreliable when researchers choose exclusion restrictions in an *ad hoc* fashion or choose none at all. To improve the implementation of selection models in accounting research, we recommend careful reporting of sensitivity analyses and robustness tests, which, surprisingly, are uncommon in accounting studies that use selection models. The majority of the 75 studies in our review do not report whether their inferences are sensitive to alternative exclusion restrictions, nor do they discuss the problems that can arise when using the selection model, such as high multicollinearity. Our central conclusion is that, as accounting researchers, we need to be more careful and rigorous in our implementation of selection models, particularly in the choice of exclusion restrictions. Further, because of the inherent limitations and fragility of selection models, we should also be much more circumspect with respect to claims about “controlling for selection bias.” Last, it may not be feasible to implement a convincing selection model in some research settings and, in this case, our advice is that studies acknowledge this limitation and provide a caveat that the reported results could be affected by selection bias.

The remainder of our article proceeds as follows. The next section discusses the selection model and implementation issues. Section III reviews how selection models have been used in the accounting literature and compares this with best practice. We also highlight the differences between our critique of selection models and those of Larcker and Rusticus (2010) and Tucker (2010), who survey the accounting literature’s implementation of regular instrumental variable (IV) estimation and Heckman models. Section IV provides three empirical examples based on past accounting studies and shows that inferences are extremely sensitive to fairly minor changes in the selection model’s specification.¹ Section V replicates and extends a study that was recently published in one of the top-tier accounting journals, demonstrating that its inferences are sensitive to minor changes in model specification. Section VI offers guidance on improving the implementation of selection models. These recommendations have important implications for editors and reviewers, as well as authors. Section VII concludes.

¹ By “fairly minor” we mean the chosen research design would not necessarily arouse the suspicions of an editor or reviewer.

II. CORRECTING FOR SELECTIVITY BIAS

The Selection Model

There are two distinct applications of the selection model. The first—commonly known as a treatment effect model—is where an endogenous indicator variable (D) is included as an independent regressor. For example, a researcher might be interested in testing whether management earnings forecasts affect the cost of capital. In this case, the endogenous indicator variable (D) indicates whether the company issues an earnings forecast and the dependent variable is the cost of capital. The second application—sometimes known as a sample selection model—occurs when a regression is estimated on a subsample of observations. For example, a researcher might be interested in testing the determinants of management forecast accuracy. In this case, the dependent variable measures forecast accuracy and the regression is estimated on a subsample of companies that issue earnings forecasts. In both applications D is endogenous, raising potential concerns about bias.

The treatment effect model can be written as follows:

$$Y = \beta'X + \theta D + u, \tag{1}$$

where X is a vector of exogenous variables (including an intercept) that affect the dependent variable, Y . The choice of D is estimated using a binary choice model:

$$D^* = a'_0Z + a'_1X + v, \tag{2}$$

where $D = 1$ if $D^* \geq 0$ and $D = 0$ if $D^* < 0$. Equation (2) is usually estimated using probit, which assumes a normally distributed error term.

The error terms in Equations (1) and (2), u and v , are assumed to have a bivariate normal distribution with mean zero and covariance matrix:

$$\begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}$$

If the error terms u and v are correlated (i.e.: $\rho \neq 0$), then $E(u|D) \neq 0$ and the OLS estimate of θ in Equation (1) will be biased. The intuition underlying the Heckman procedure is to control for this bias by estimating the inverse Mills' ratio (*MILLS*) from Equation (2):

$$MILLS = E(u|D) = \begin{cases} \frac{\varphi(\hat{\alpha}'_0Z + \hat{\alpha}'_1X) / \Phi(\hat{\alpha}'_0Z + \hat{\alpha}'_1X)}{-\varphi(\hat{\alpha}'_0Z + \hat{\alpha}'_1X) / (1 - \Phi(\hat{\alpha}'_0Z + \hat{\alpha}'_1X))} & \text{if } D = 1 \\ & \text{if } D = 0, \end{cases}$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are the normal density and cumulative distribution functions, respectively. The researcher then controls for selection bias by adding *MILLS* to Equation (1), which becomes:

$$Y = \beta'X + \theta D + \rho\sigma MILLS + \varepsilon. \tag{3}$$

The error term ε in Equation (3) is uncorrelated with D , which means that θ is estimated without bias. The presence and direction of selection bias is inferred from the statistical significance and sign of the *MILLS* coefficient in Equation (3).² Note that the *MILLS* and D variables are correlated by construction, due to the fact that *MILLS* is a function of D , as defined above. Moreover, the

² Equations (2) and (3) can be estimated using the traditional two-step approach or maximum likelihood (ML).

MILLS variable is also correlated by construction with the X variables in Equation (3). As we illustrate later, these correlations can result in high levels of multicollinearity.

The sample selection model is conceptually the same as the treatment effect model except that Equation (3) is estimated for a subsample of observations rather than the full sample. For example, an accounting researcher may wish to estimate a model of forecast accuracy (Y) on a subsample of companies that issue forecasts ($D = 1$). In this case, Equation (3) becomes $Y = \beta'X + \theta + \rho\sigma\text{MILLS}_1 + \varepsilon$, where:

$$\text{MILLS}_1 = \varphi(\hat{\alpha}'_0Z + \hat{\alpha}'_1X) / \Phi(\hat{\alpha}'_0Z + \hat{\alpha}'_1X).$$

Alternatively, accounting researchers may estimate separate Y models for the subsamples where $D = 1$ and $D = 0$. For example, [Chaney et al. \(2004\)](#) estimate separate audit fee models for companies that hire Big N auditors ($D = 1$) and companies that hire non-Big N auditors ($D = 0$). When Equation (3) is estimated on the $D = 0$ subsample, the Y model becomes $Y = \beta'X + \rho\sigma\text{MILLS}_0 + \varepsilon$, where:

$$\text{MILLS}_0 = -\varphi(\hat{\alpha}'_0Z + \hat{\alpha}'_1X) / (1 - \Phi(\hat{\alpha}'_0Z + \hat{\alpha}'_1X)).$$

Implementation Issues

The difference between the OLS model in Equation (1) and the selection model in Equation (3) is that the latter includes *MILLS* as an additional independent variable. Identification of selection bias comes from two sources:

- (1) *MILLS* is nonlinear in its arguments (the X and Z variables), and
- (2) the Z variables are excluded from Equation (3).

The Z variables are also known as “exclusion restrictions” because the researcher assumes they have no direct impact on Y , such that any association between Y and Z is indirect through the *MILLS* variable.

It is well known in econometrics that the researcher’s choice of exclusion restrictions is vital for implementing the selection model in a way that convincingly controls for endogeneity in D ([Little 1985](#); [Little and Rubin 1987](#); [Manning et al. 1987](#)). First, the Z variables must be exogenous, otherwise the first stage coefficient estimates (and therefore *MILLS*) will be biased. Second, the Z variables need to be important determinants of D in Equation (2) in order for the *MILLS* variable to yield a powerful test for selection bias in Equation (3). Finally, it must be valid to exclude the Z variables from Equation (3). If a Z variable is incorrectly omitted from the Y model, then there is a classical correlated omitted variable problem. That is, a relevant Z variable is omitted from the second stage Y model and is correlated with the *MILLS*—which is a function of Z , X , and D —causing the *MILLS* coefficient to be biased. In turn, this means that *MILLS* will not properly control for the endogenous component of the D variable, such that the exogenous effect of D in Equation (3) will be estimated with bias.

In many applications, the difficulty lies in finding good Z variables. While not good practice, it is possible to estimate selection models with no exclusion restrictions (i.e., no Z ’s). The *MILLS* coefficients are technically identifiable even if the researcher imposes no exclusion restrictions because *MILLS* is nonlinear in its arguments. However, econometricians recommend against using nonlinearities to identify selection bias for two reasons ([Little 1985](#)). First, without Z variables, identification of the selection bias comes solely from the (untested) functional form assumptions. To illustrate, if Y is actually a nonlinear function of the X variables but the researcher incorrectly assumes that the relation is linear, then the *MILLS* variable will capture this functional form misspecification. The problem is that the researcher typically does not know whether the

independent variables affect Y in a linear or nonlinear way and functional form assumptions typically have no firm basis in theory.

Second, the selection model is more likely to suffer from multicollinearity problems when there are no exclusion restrictions. Multicollinearity can arise in the selection model because, by construction, $MILLS$ is correlated with the independent variables in the second stage (i.e., X and D in Equation (3)). Multicollinearity is more likely to be problematic when there are no exclusion restrictions (i.e., no Z variables) because $MILLS$ is close to being linear over a wide range of its values (Manning et al. 1987; Puhani 2000; Li and Prabhala 2007).

There are two issues raised by having high multicollinearity. First, the coefficient standard errors are inflated, making it less likely that the coefficient estimates are statistically significant. We demonstrate in Section IV that inflated standard errors are important because they can result in the $MILLS$ coefficient becoming statistically insignificant in the second stage estimation, which can lead to an incorrect inference with respect to the need to control for selection bias. In particular, we show that—even when the $MILLS$ coefficient is statistically insignificant—inferences from the selection model can be different from those obtained without including $MILLS$. In this situation, the selection model indicates that it is legitimate to omit the insignificant $MILLS$ variable, and yet, omitting $MILLS$ gives different inferences for the treatment variable (D) because multicollinearity is then much lower.

A second problem arises in a model that is not correctly specified. It is well known that when the model is correctly specified, the coefficients are estimated without bias even when multicollinearity is high. However, a less well-known problem arises when the model is misspecified, because in that situation multicollinearity can exacerbate the bias (Thursby 1988). The risk of misspecification is particularly high in the selection model because the statistical identification of selectivity is through the assumed exclusion restrictions and the assumed functional form. If the researcher's assumptions are incorrect, then the model is misspecified and, together with high multicollinearity, this can result in large biases. This is illustrated by our empirical examples in Section IV, where we show that relatively minor changes to the model specification can reverse the signs on the coefficients of interest and these opposing coefficients can attain high levels of statistical significance even when multicollinearity is high.

III. SURVEY OF SELECTION MODELS IN ACCOUNTING RESEARCH

This section evaluates whether the implementation issues discussed in Section II are relevant to the accounting literature. We begin by searching *The Accounting Review*, *Journal of Accounting and Economics*, *Journal of Accounting Research*, *Review of Accounting Studies*, and *Contemporary Accounting Research* for articles that use selection models. The search is undertaken electronically using the keywords “endogeneity,” “Heckman,” “selection,” and “treatment effect,” and results in 75 articles in the ten-year period from 2000–2009. Topic areas are shown in Table 1, with 16 articles in auditing, 16 in disclosure, 13 in earnings management/quality, 11 in contracting/corporate governance, 2 in tax, 16 in other financial accounting topics, and 1 in management accounting.

Panel A of Table 2 shows a sharp increase in selection models, with 50 of the 75 articles (67 percent) published in the most recent four-year period 2006–2009, compared with just 25 (33 percent) in the earlier five-year period (2000–2005). Panel B shows that 52 studies (69 percent) estimate treatment effect models, while 23 studies (31 percent) use the selection model to control for the fact that they are estimating the model on a non-random subsample. There are 32 studies (43 percent) that use selection models for their primary analysis. The remaining 43 studies (57 percent) use selection models as a secondary analysis to corroborate inferences that are obtained without controlling for selection.

Table 2, Panel C reports that 54 studies (72 percent) follow the recommended econometric practice of imposing one or more exclusion restrictions in the second stage model. However, a

TABLE 1
Accounting Studies that Use Selection Models
(2000–2009)

Disclosure	Auditing	Other Financial Accounting
Leuz and Verrecchia (2000)	Kim et al. (2003)	Beatty et al. (2002)
Bushee et al. (2003)	Weber and Willenborg (2003)	Lang et al. (2003)
Leuz (2003)	Chaney et al. (2004)	Ball and Shivakumar (2005)
Baginski et al. (2004)	Johnstone et al. (2004)	Cowen et al. (2006)
Atiase et al. (2005)	Khurana and Raman (2004)	Garfinkel and Sokobin (2006)
Ahmed et al. (2006)	Mansi et al. (2004)	Ke and Yu (2006)
Baber et al. (2006)	Pittman and Fortin (2004)	Bens and Johnston (2009)
Marques (2006)	Louis (2005)	Butler et al. (2007)
Verrecchia and Weber (2006)	Guedhami and Pittman (2006)	DeFond and Hung (2007)
Engel et al. (2007)	Fortin and Pittman (2007)	Eberhart et al. (2008)
Tucker (2007)	Ruddock et al. (2007)	Blouin and Krull (2009)
Baber and Gore (2008)	Behn et al. (2008)	Feng et al. (2009)
Rogers (2008)	Choi et al. (2008)	Jackson et al. (2009)
Christensen et al. (2009)	Chang et al. (2009)	Kini et al. (2009)
Collins et al. (2009)	Choi et al. (2009)	Wier (2009)
Hui et al. (2009)	Li (2009)	Zhang (2009)
Earnings Management/Quality	Contracting/Corporate Governance	Tax
Barton (2001)	Roulstone (2003)	Phillips (2003)
Pincus and Rajgopal (2002)	Fee and Hadlock (2004)	Omer et al. (2006)
Schrand and Wong (2003)	Asquith et al. (2005)	
Haw et al. (2004)	Davila and Penalva (2006)	Management Accounting
Kim and Yi (2006)	Blacconiere et al. (2008)	Muller and Riedl (2002)
Morsfield and Tan (2006)	Cheng and Farber (2008)	
Doyle et al. (2007)	Frankel et al. (2008)	
Chung and Wynn (2008)	Carter et al. (2009)	
Jiang (2008)	Hoitash et al. (2009)	
Louis et al. (2008)	Lara et al. (2009)	
Files et al. (2009)	Nichols et al. (2009)	
Gong et al. (2009)		
Katz (2009)		

The sample is based on a search for the terms “endogeneity,” “Heckman,” “selection,” and “treatment effect” for articles in *The Accounting Review*, *Journal of Accounting and Economics*, *Journal of Accounting Research*, *Review of Accounting Studies*, and *Contemporary Accounting Research*.

surprising number of studies report selection models without exclusion restrictions. Eight studies have no Z variables whatsoever, while six report selection models both with and without exclusion restrictions. In another seven studies, the authors do not report the independent variables in the first stage models, so we are unable to determine whether they impose exclusion restrictions and it is impossible to evaluate the suitability of such restrictions. In summary, between 19 and 28 percent of accounting studies report selection models without exclusion restrictions. Of the 60 articles that do impose exclusion restrictions, only three report whether their results are robust to alternative restrictions. Moreover, only three of the 60 articles that employ Z variables attempt to provide an economic or theoretical rationale for the exclusions. This is important because, as shown later, *ad hoc* exclusion restrictions can yield non-robust inferences.

TABLE 2
Descriptive Statistics for 75 Accounting Studies that Use Selection Models (2000–2009)

Panel A: The Increasing Use of Selection Models in the Accounting Literature

	<u>2000</u>	<u>2001</u>	<u>2002</u>	<u>2003</u>	<u>2004</u>	<u>2005</u>	<u>2006</u>	<u>2007</u>	<u>2008</u>	<u>2009</u>	<u>Total</u>
Number of studies per year	1	1	3	8	8	4	12	7	11	20	75

Panel B: Purpose of the Selection Model

Do studies estimate a treatment effects model or a sample selection model?	Treatment Effects 52 (69%)	Sample Selection 23 (31%)	Total 75 (100%)
Do studies use the selection model for their primary analysis or as a secondary robustness test?	Primary 32 (43%)	Secondary 43 (57%)	Total 75 (100%)

Panel C: Implementation Issues

Do studies impose one or more exclusion restrictions in the second stage model?	Yes 54 (72%)	Yes/No ^a 6 (8%)	No 8 (11%)	Unclear ^b 7 (9%)	Total 75 (100%)
Of the 60 studies that impose exclusion restrictions, do the studies report robustness results using different restrictions?	Yes 3 (5%)		No 57 (95%)		Total 60 (100%)
Of the 60 studies that impose exclusion restrictions, do the studies attempt to justify them on economic grounds?	Yes 3 (5%)		No 57 (95%)		Total 60 (100%)
Do studies report multicollinearity diagnostics for the endogenized regressor and the inverse Mills' ratio?	Yes 3 (4%)		No 72 (96%)		Total 75 (100%)

^a The “Yes/No” group refers to studies that report alternate specifications both with and without an exclusion restriction.
^b The “Unclear” group refers to studies where it is not disclosed whether any exclusion restrictions are imposed in the second stage. This situation arises when studies do not report the set of independent variables in the first stage model. The sample is based on a search for the terms “endogeneity,” “Heckman,” “selection,” “treatment effects” for articles in *The Accounting Review*, *Journal of Accounting and Economics*, *Journal of Accounting Research*, *Review of Accounting Studies*, and *Contemporary Accounting Research*.

Studies typically rely on the existing literature to justify their choice of independent variables in the first and second stage model specifications. However, the critical issue for the selection model is the researcher’s choice of *exclusion* restrictions, that is, which of the independent variable(s) in the first stage model should be excluded from the second stage model. In this respect, many accounting studies fail to justify their exclusion of *Z* variables from the second stage. Some studies do explicitly point out that exclusion restrictions are desirable from an econometric point of view, but they nevertheless fail to explain why their chosen exclusion restrictions are valid. In most cases, the exclusion restrictions appear to be chosen in an *ad hoc* way without obvious justification. Likewise, Larcker and Rusticus (2010) (hereafter, LR) report that accounting studies using regular

instrumental variables (IV) typically fail to justify their exclusion restrictions.³ Finally, although selection models can suffer from high levels of multicollinearity (Manning et al. 1987; Puhani 2000), only three of the 75 studies report formal tests of multicollinearity.

We conclude that many accounting studies implement the selection model in a mechanical way with limited appreciation of the econometric issues surrounding its implementation. This appraisal is similar to LR's survey, but there are important differences between their critique and ours. The endogenous variables are *continuous* in the regular IV procedures examined by LR, whereas there is an endogenous *indicator* variable (i.e., D) in the selection model. This means that the correction for endogeneity is different in the selection model compared with regular IV. In particular, the selection model uses the *MILLS* variable to control for the correlation between error terms, but there is no equivalent of *MILLS* in regular IV. This makes our critique different from LR in two important ways. First, *MILLS* is nonlinear in its arguments, which means that the selection model can technically be estimated even in the absence of any exclusion restrictions. In contrast, at least one exclusion restriction is required for identification in the regular IV approach. In the following section, we demonstrate that the selection model can yield non-robust inferences and multicollinearity is often high when the researcher chooses no exclusion restrictions.

The second difference between regular IV and selection models is that researchers can use the statistical significance of the *MILLS* coefficient to assess the presence or absence of selection bias. However, as discussed in Section II, multicollinearity can complicate these inferences. In the following section, we shall demonstrate that—even when the *MILLS* coefficient is statistically insignificant, suggesting no selection bias—key conclusions reached from the selection model can still be very different from OLS. A lack of statistical significance can be caused by high multicollinearity, so a statistically insignificant *MILLS* coefficient does not necessarily mean there is no selection problem. Further, even when the *MILLS* coefficient is statistically significant, we demonstrate that inferences from the selection model are likely to be unreliable if exclusion restrictions are absent or are chosen in an *ad hoc* manner.

A final difference between our study and LR is that we apply our critique to a study that has been published in a top-tier accounting journal (Section V). This is important because the empirical examples in Section IV and in LR are homemade in the sense that they do not directly replicate a published study. A skeptic might argue that our examples, as well as the example in LR, are extreme and not representative of the implementation of the selection model in published studies that pass the scrutiny of experienced editors and reviewers in top-tier journals. We show that this is not the case by illustrating implementation problems and non-robustness using a published study.

Our study is also related to a discussion by Tucker (2010) about the relative merits of propensity score matching (PSM) versus the inverse Mills' ratio (IMR) approach. Tucker (2010) points out that these two approaches are not substitutes for each other because they are designed to address different problems. In particular, the PSM methodology controls for selection on "observables" i.e., it controls for the correlation between the error term in the first stage model (u) and the independent variables in the second stage (i.e., X and Z). In contrast, the IMR approach controls for selection on "unobservables" (i.e., it controls for the correlation between u and v).⁴ Our

³ Our study is also related to Li and Prabhala (2007) who survey various applications of the selection model in the corporate finance literature. However, unlike our study, they do not offer a viewpoint as to whether researchers have applied the techniques correctly and they do not illustrate the fragility of the selection model using empirical examples.

⁴ Tucker (2010) makes two other points about Heckman models in the accounting literature. First, she points out that many studies fail to report which formulas they use to calculate the IMR variables and this lack of disclosure makes it difficult for the reader to interpret the signs of the coefficients on the IMR variables. Second, she notes that the two-step procedure is less efficient than maximum likelihood (ML) estimation. On this second point, we caution the reader that ML is not necessarily preferable despite its greater efficiency, as ML can yield inferences that are less robust than the two-step procedure (Puhani 2000).

study is different from [Tucker \(2010\)](#) because we emphasize the problems that arise when the Heckman model is estimated without any exclusion restrictions or when the exclusion restrictions are chosen in an *ad hoc* manner. Unlike [Tucker \(2010\)](#), we illustrate the problems of fragility and high multicollinearity using empirical examples taken from the accounting literature.

IV. THREE EMPIRICAL EXAMPLES

We have seen that accounting studies typically do not justify the exclusion restrictions that appear in the second stage outcome model and in some instances they fail to impose any exclusion restrictions at all. Our aim in this section is to demonstrate that these practices can yield results that are far too sensitive to provide reliable and robust inferences. We obtain data for the empirical analyses in this section from Compustat, Audit Analytics, CRSP, and I/B/E/S for the period 2000–2009.

Auditing studies often employ selection models to control for endogeneity in the company's selection of a Big N or non-Big N audit firm ([Kim et al. 2003](#); [Weber and Willenborg 2003](#); [Chaney et al. 2004](#); [Khurana and Raman 2004](#); [Mansi et al. 2004](#); [Louis 2005](#); [Fortin and Pittman 2007](#); [Behn et al. 2008](#); [Choi et al. 2008](#); [Li 2009](#)). In fact, auditor choice is the most common first stage model in our survey, accounting for ten studies (13.3 percent). We catalogue a wide divergence in the independent variables used to estimate the first stage auditor-choice models in these articles. For example, six studies control for company size using the log of total assets ([Weber and Willenborg 2003](#); [Chaney et al. 2004](#); [Mansi et al. 2004](#); [Louis 2005](#); [Fortin and Pittman 2007](#); [Choi et al. 2008](#)); three studies use the log of sales rather than assets ([Kim et al. 2003](#); [Khurana and Raman 2004](#); [Li 2009](#)); and one study employs the log of market value ([Behn et al. 2008](#)). The studies also differ in how they control for company profitability: three studies use an indicator variable for losses ([Khurana and Raman 2004](#); [Choi et al. 2008](#); [Li 2009](#)); two use a continuous measure of profitability ([Mansi et al. 2004](#); [Louis 2005](#)); and three incorporate both a loss indicator variable and a continuous profitability variable ([Chaney et al. 2004](#); [Fortin and Pittman 2007](#); [Behn et al. 2008](#)). Our point here is not to say that studies should use identical measures of company size and profitability. Rather, our point is that the different exclusion restrictions that studies have imposed on these variables can have a dramatic impact on inferences, as we illustrate below.

The ten studies also differ in their use of exclusion restrictions. Two studies fail to impose any exclusion restrictions, as all the independent variables from the first stage auditor-choice model are employed as regressors in the second stage outcome model ([Chaney et al. 2004](#); [Li 2009](#)). These studies estimate the effects of selection bias through the nonlinearity in the inverse Mills' ratio even though the econometrics literature advises against this practice.

Of the eight studies that impose exclusion restrictions, none provide an explicit explanation or economic rationale for why the restrictions are valid. For example, [Kim et al. \(2003\)](#) include the log of sales and a loss indicator variable in the first stage model of auditor choice, but these two variables are excluded from their second stage model of discretionary accruals. [Weber and Willenborg \(2003\)](#) include the log of assets in their first stage auditor-choice model. They drop the log of assets and use an alternative size measure, the log of sales, in their second stage audit-opinion model. [Khurana and Raman \(2004\)](#) include the log of sales and an indicator variable for losses in their auditor-choice model, but these variables are not included in their second stage model of the cost of equity. [Mansi et al. \(2004\)](#) control for company size and profitability in their auditor-choice model, but these variables are excluded in their second stage model of the cost of debt. [Fortin and Pittman \(2004\)](#) include company size, a loss indicator variable, and a continuous profitability variable in their first stage auditor-choice model, but these variables are excluded in their second stage models of bond ratings and yield spreads. [Behn et al. \(2008\)](#) control for both profitability and the occurrence of losses in their auditor-choice model, but they exclude the continuous profitability

variable from their second stage model of analyst forecast accuracy. Finally, Choi et al. (2008) include a cross-country measure of legal liability in their auditor-choice model, but they assume that legal liability has no impact on audit fees in the second stage. In summary, the exclusion restrictions in these studies appear to be chosen in an *ad hoc* way, the restrictions are not justified or explained, and they lack any obvious economic rationale.

The objective of this discussion is not to criticize these audit studies, but to illustrate the problems that can arise when researchers have either no exclusion restrictions or impose arbitrary restrictions. To illustrate the former, we include in the second stage model all the independent variables from the first stage auditor-choice model. To demonstrate the latter, we exclude from the second stage models one of the variables for company size or company profitability, since the prior studies cited above commonly impose exclusion restrictions on these two variables.

Our first stage model of auditor choice employs a set of independent variables that are based on past studies. As well as controlling for company size, we include an indicator variable for losses ($Loss_{it}$) and a continuous measure of profitability (ROA_{it}), the ratio of total liabilities to total assets ($Leverage_{it}$), the current ratio ($Liquidity_{it}$), external financing (New_Issue_{it}), and the lag between the fiscal year-end and the audit report date ($Report_Lag_{it}$). Table 3 reports the results for model 1, which controls for company size using the log of total assets ($LnTA_{it}$), and model 2, which uses the log of sales ($LnSale_{it}$) to control for size.

The dependent variables in the second stage are taken from the prior audit literature that uses selection models to examine the relation between auditor choice and: (1) audit fees, (2) the cost of equity, and (3) going-concern audit opinions. We start by examining the relation between audit fees and auditor choice in Panel A of Table 4. The $MILLS_{it}$ in Columns (1) and (2) is calculated from model 1 of Table 3. Columns (3) and (4) use the $MILLS_{it}$ calculated from model 2 of Table 3. In Columns (1), (4), and (5), we control for company size using the log of assets ($LnTA_{it}$). In Columns (2), (3), and (6), we control for company size using the log of sales ($LnSale_{it}$). There are no exclusion restrictions in Columns (1) and (3) because the same company size measure is used in both the first and second stage models. In Column (2) there is an exclusion restriction on $LnTA_{it}$ because $LnTA_{it}$ is included in the first stage but not in the second stage (instead $LnSale_{it}$ is used in the second stage). In Column (4) there is an exclusion restriction on $LnSale_{it}$ because $LnSale_{it}$ is included in the first stage but not in the second stage (instead $LnTA_{it}$ is used in the second stage). In each second stage model we include year indicators and additional control variables ($Loss_{it}$, ROA_{it} , $Busy_{it}$, $Foreign_{it}$, $Special_Item_{it}$, $New_Auditor_{it}$, $Leverage_{it}$, $Liquidity_{it}$, $Invrec_{it}$, New_Issue_{it} , $Report_Lag_{it}$) that are defined in the table notes. Columns (5) and (6) are the OLS models that do not control for selection bias (i.e., they do not include the $MILLS_{it}$ variables). Except for the exclusion of $MILLS_{it}$, the model specification in Column (5) is identical in Columns (1) and (4) as company size is controlled for using the log of assets ($LnTA_{it}$). Likewise, Column (6) is the same as Columns (2) and (3) because company size is controlled for using the log of assets ($LnTA_{it}$), but Column (6) does not control for selection bias. Thus, Columns (5) and (6) are the same as Columns (1) to (4), with the exception that they do not include the $MILLS_{it}$ variables. This means that we are able to infer whether any fragility in inferences and high multicollinearity in the second stage selection models are due to the inclusion of the control for selection bias.

The results in Columns (1), (2), and (4) show that the Big_{it} coefficients are significantly positive, indicating that Big 5 audit firms earn fee premiums. However, the exact opposite is obtained in Column (3), where the Big_{it} coefficient is significantly negative. Thus, the findings for the treatment variable (Big_{it}), are sensitive to the choice of exclusion restriction in the second stage. In contrast, the OLS results do not display the same fragility, as the Big_{it} coefficients are significantly positive regardless of whether size is controlled for using $LnSale_{it}$ or $LnTA_{it}$. Moreover, the coefficient magnitudes are quite close in the OLS models (0.369 and 0.492), whereas they are markedly different across the four specifications of the selection model (0.325, 1.295, -0.333, and 0.774).

TABLE 3
First Stage Models of Auditor Choice

	<u>Model 1</u>	<u>Model 2</u>
$LnTA_{it}$	0.587*** (45.09)	
$LnSale_{it}$		0.428*** (44.89)
$Loss_{it}$	0.211*** (7.98)	0.248*** (10.02)
ROA_{it}	-0.205*** (-14.18)	-0.127*** (-9.85)
$Leverage_{it}$	-0.232*** (-5.45)	-0.220*** (-6.43)
$Liquidity_{it}$	0.026*** (5.74)	0.072*** (13.70)
New_Issue_{it}	-0.169*** (-7.46)	0.019 (0.90)
$Report_Lag_{it}$	0.000 (0.32)	-0.000 (-0.86)
Intercept $_{it}$	-2.759*** (-34.47)	-1.990*** (-28.70)
Observations	49,548	49,548
Pseudo-R ²	42.94	34.09
Year Variables	Yes	Yes

*** Significant at the 1 percent level (two-tailed).

The continuous variables are winsorized at the 1 percent and 99 percent percentiles in order to control for outliers. The standard errors are adjusted for time-series dependence by clustering on each company (Petersen 2009). All the models include year indicator variables but the year effects are not tabulated.

The dependent variable (Big_{it}) equals 1 if the company's auditor is a Big 5 firm, and 0 otherwise.

Variable Definitions:

$LnTA_{it}$ = natural log of total assets;

$LnSale_{it}$ = natural log of sales;

$Loss_{it}$ = indicator variable equal to 1 if the previous year's net income is negative, and 0 otherwise;

ROA_{it} = income before extraordinary items scaled by total assets;

$Leverage_{it}$ = total liabilities scaled by total assets;

$Liquidity_{it}$ = total current assets scaled by total current liabilities;

New_Issue_{it} = indicator variable equal to 1 if a company's total shares outstanding increase by 5 percent, or total debts increase by 5 percent relative to total assets, and 0 otherwise; and

$Report_Lag_{it}$ = number of days between the fiscal year-end and the signing date of auditors' reports.

Not only are the results fragile for the endogenous regressor (Big_{it}), but they are also sensitive for $MILLS_{it}$. The $MILLS_{it}$ coefficients are significantly negative in Columns (2) and (4), significantly positive in Column (3), and insignificant in Column (1). Thus, any conclusions about the existence and direction of selection bias are wholly dependent on the researcher's choice of exclusion restrictions. In particular, it is easy to obtain contrary results for the Big_{it} and $MILLS_{it}$ variables by choosing either no exclusion restrictions or arbitrary restrictions for the company size variables. And, as illustrated by our survey of the literature, these research design choices are

TABLE 4

The Second Stage Treatment Effect Models and Models that Do Not Control for Endogeneity

Panel A: The Relation between Audit Fees and Auditor Size

	(1)	(2)	(3)	(4)	(5)	(6)
	Second Stage Models				OLS Models	
<i>Big_{it}</i>	0.325*** (6.28)	1.295*** (25.84)	-0.333*** (-4.00)	0.774*** (17.22)	0.369*** (22.70)	0.492*** (27.18)
<i>MILLS_{it}</i>	0.028 (0.89)	-0.539*** (-17.88)	0.500*** (9.73)	-0.252*** (-9.66)		
Variance-Inflation-Factors (VIFs)						
<i>Big_{it}</i>	13.33	9.49	25.17	12.29	1.70	1.66
<i>MILLS_{it}</i>	7.97	5.83	16.47	7.85		

*** Significant at the 1 percent level (two-tailed).

The continuous variables are winsorized at the 1 percent and 99 percent percentiles in order to control for outliers. The standard errors are adjusted for time-series dependence by clustering on each company (Petersen 2009). Z- and t-statistics are reported in parentheses.

The dependent variable is the log of audit fees ($Ln(audit\ fees_{it})$). The treatment variable (Big_{it}) equals 1 if the company's auditor is a Big 5 firm, and 0 otherwise. The inverse Mills' ratios ($MILLS_{it}$) in Columns (1) and (2) are calculated from Model 1 of Table 3. Columns (3) and (4) use the $MILLS_{it}$ calculated from Model 2 of Table 3. In Columns (1), (4), and (5), we control for company size using the log of total assets ($LnTA_{it}$). In Columns (2), (3), and (6), we control for company size using the log of sales ($LnSale_{it}$). There are no exclusion restrictions in Columns (1) and (3) because the same company size measure is used in both the first stage and second stage models. In Column (2) there is an exclusion restriction on $LnTA_{it}$ because $LnTA_{it}$ is included in the first stage but not in the second stage. In Column (4) there is an exclusion restriction on $LnSale_{it}$ because $LnSale_{it}$ is included in the first stage but not in the second stage. In each model we include year indicators and additional control variables ($Loss_{it}$, ROA_{it} , $Busy_{it}$, $Foreign_{it}$, $Special_Item_{it}$, $New_Auditor_{it}$, $Leverage_{it}$, $Liquidity_{it}$, $Invrec_{it}$, New_Issue_{it} , $Report_Lag_{it}$) that are defined in the table notes. Results for the control variables are available from the authors.

Panel B: The Relation between the Cost of Equity and Auditor Size

	(1)	(2)	(3)	(4)	(5)	(6)
	Second Stage Models				OLS Models	
<i>Big_{it}</i>	-0.016*** (-4.55)	-0.035*** (-9.37)	-0.012*** (-2.80)	0.009** (2.21)	-0.005*** (-4.86)	-0.007*** (-6.42)
<i>MILLS_{it}</i>	0.006*** (3.42)	0.016*** (7.99)	0.002 (1.21)	-0.008*** (-3.63)		
Variance-Inflation-Factors (VIFs)						
<i>Big_{it}</i>	14.10	13.06	17.26	14.48	1.12	1.11
<i>MILLS_{it}</i>	12.66	11.78	15.59	13.03		

, * Significant at the 5 percent and 1 percent levels, respectively (two-tailed).

The continuous variables are winsorized at the 1 percent and 99 percent percentiles in order to control for outliers. The standard errors are adjusted for time-series dependence by clustering on each company (Petersen 2009). Z- and t-statistics are reported in parentheses.

The dependent variable is the cost of equity (COE_{it}). The treatment variable (Big_{it}) equals 1 if the company's auditor is a Big 5 firm, and 0 otherwise. The $MILLS_{it}$ in Columns (1) and (2) are calculated from Model 1 of Table 3. Columns (3) and (4) use the $MILLS_{it}$ calculated from model 2 of Table 3. In Columns (1), (4), and (5), we control for company size using the log of total assets ($LnTA_{it}$). In Columns (2), (3), and (6), we control for company size using the log of sales ($LnSale_{it}$). There are no exclusion restrictions in Columns (1) and (3) because the same company size measure is used in both the first stage and second stage models. In Column (2) there is an exclusion restriction on $LnTA_{it}$ because $LnTA_{it}$ is included in the first stage but not in the second stage. In Column (4) there is an exclusion restriction on $LnSale_{it}$ because $LnSale_{it}$ is included in the first stage but not in the second stage. In each model we include year indicators and additional control variables ($Loss_{it}$, ROA_{it} , $LnMB_{it}$, $Growth_{it}$, $Beta_{it}$, $Volatility_{it}$, $Leverage_{it}$, $Liquidity_{it}$, New_Issue_{it} , $Report_Lag_{it}$) that are defined in the table notes. Results for the control variables are available from the authors.

(continued on next page)

TABLE 4 (continued)

Panel C: The Relation between Going-Concern Audit Opinions and Auditor Size

	(1) Second Stage Models	(2)	(3) Ordinary Probit Models	(4)
Big_{it}	-1.295*** (-8.05)	0.014 (0.08)	-0.291*** (-9.45)	-0.230*** (-7.70)
$MILLS_{it}$	0.590*** (6.19)	-0.144 (-1.46)		
Variance-Inflation-Factors (VIFs)				
Big_{it}	25.97	24.39	1.58	1.57
$MILLS_{it}$	16.62	15.68		

*** Significant at the 1 percent level (two-tailed).

The continuous variables are winsorized at the 1 percent and 99 percent percentiles in order to control for outliers. The standard errors are adjusted for time-series dependence by clustering on each company (Peterson 2009). Z- and t-statistics are reported in parentheses.

The dependent variable (GC_{it}) equals 1 if company i receives a going-concern opinion in year t , and 0 otherwise. The treatment variable (Big_{it}) equals 1 if the company's auditor is a Big 5 firm, and 0 otherwise. The $MILLS_{it}$ in Columns (1) and (2) are calculated from model 2 of Table 3. In Columns (1) and (3), we control for profitability using both an indicator variable for losses ($Loss_{it}$), and a continuous measure of the return on assets (ROA_{it}). In Columns (2) and (4), we control for profitability using just the return on assets (ROA_{it}). In every model, we control for company size using the log of sales ($LnSale_{it}$). There are no exclusion restrictions in Column (1) because the $Loss_{it}$ and ROA_{it} variables are included in both the first stage and second stage models. There is an exclusion restriction in Column (2) because the $Loss_{it}$ variable is used in the first stage model but not in the second stage model. In each model, we include year indicators and additional control variables ($New_Auditor_{it}$, $Leverage_{it}$, $Liquidity_{it}$, New_Issue_{it} , $Report_Lag_{it}$) that are defined in the table notes. Results for the control variables are available from the authors.

Variable Definitions:

Big_{it} = 1 if the company's auditor is a Big 5 firm, and 0 otherwise;

$MILLS_{it}$ = inverse Mills' ratio calculated from the auditor-choice models in Table 3;

$Ln(audit\ fees_{it})$ = natural log of audit fees;

$LnTA_{it}$ = natural log of total assets;

$LnSale_{it}$ = natural log of sales;

$Loss_{it}$ = indicator variable equal to 1 if the previous year's net income is negative, and 0 otherwise;

ROA_{it} = income before extraordinary items scaled by total assets;

$Busy_{it}$ = indicator variable equal to 1 if the company's fiscal year-end is from December to March, and 0 otherwise;

$Foreign_{it}$ = indicator variable equal to 1 if a company has foreign operations, and 0 otherwise;

$Special_Item_{it}$ = special items scaled by total assets;

$New_Auditor_{it}$ = indicator variable equal to 1 if a sample year is the first year when a new auditor is appointed, and 0 otherwise;

$Leverage_{it}$ = total liabilities scaled by total assets;

$Liquidity_{it}$ = total current assets scaled by total current liabilities;

$Invrec_{it}$ = inventories plus account receivables, scaled by total assets;

New_Issue_{it} = indicator variable equal to 1 if a company's total shares outstanding increase by 5 percent, or total debts increase by 5 percent relative to total assets, and 0 otherwise;

$Report_Lag_{it}$ = number of days between the fiscal year-end and the signing date of the audit report;

COE_{it} = average cost of equity in the fiscal year-end month (the measure is constructed following Gebhardt et al. [2001], Claus and Thomas [2001], Gode and Mohanram [2003], and Easton [2004]);

$Beta_{it}$ = stock's beta in each company year (obtained by regressing the stock's daily returns against the corresponding daily returns on the CRSP value-weighted index);

$Volatility_{it}$ = variance of the residuals from regressing daily returns for the fiscal year against the daily CRSP value-weighted index;

$LnMB_{it}$ = natural logarithm of the market-to-book ratio; and

$Growth_{it}$ = long-term growth rate of analysts' earnings forecasts (if the long-term growth rate is not available, then we instead take the ratio of analysts' three-year-ahead earnings forecasts to the two-year-ahead earnings forecasts minus 1).

common in past auditor choice studies. That is, some studies impose no exclusion restrictions, some exclude the log of sales from the second stage, and others exclude the log of assets.

We also find high multicollinearity in the second stage models. For example, in Panel A, the VIFs range from 9.49 to 25.17 for Big_{it} and from 5.83 to 16.47 for $MILLS_{it}$.⁵ In contrast, the VIFs are low in the OLS models. Thus, multicollinearity is high due to the inclusion of $MILLS_{it}$ in the selection model.

We employ the same exclusion restrictions on the two company size variables in Panel B for our analysis of the cost of equity. Again, results are fragile in the selection models, with significant negative coefficients on Big_{it} in Columns (1), (2), and (3), and a significant positive coefficient on Big_{it} in Column (4). In contrast, the OLS results are robust, as the Big_{it} coefficients have the same negative sign, they are both statistically significant, and the coefficient magnitudes are similar (−0.005 and −0.007). Inferences in the second stage models are also shaky with respect to the tests for selection bias. The $MILLS_{it}$ coefficients are significantly positive in Columns (1) and (2), insignificant in Column (3), and significantly negative in Column (4). Again, the VIFs are generally high in the selection models, whereas they are low in the OLS models.⁶

To demonstrate that researchers need not only be concerned about exclusion restrictions on company size variables, in the next example we show that inferences can also be sensitive to exclusion restrictions on variables that measure profitability. This analysis is again motivated by the evidence in our survey of the auditor choice literature: studies employ different measures of profitability—e.g., a loss indicator variable ($Loss_{it}$) or the return on assets (ROA_{it})—and utilize different exclusion restrictions on these variables.

In Panel C of Table 4, we explore the relation between going-concern (GC) audit opinions and auditor choice. The $MILLS_{it}$ variable in Columns (1) and (2) is calculated from model 2 of Table 3, which controls for company size using the log of sales ($LnSale_{it}$). In Columns (1) and (3), we include both an indicator variable for losses ($Loss_{it}$) and a continuous measure of the return on assets (ROA_{it}), whereas in Columns (2) and (4) we exclude the $Loss_{it}$ variable. There are no exclusion restrictions in Column (1) because the $Loss_{it}$ and ROA_{it} variables are included in both the first and second stage models. There is an exclusion restriction in Column (2) because, although we control for the return on assets (ROA_{it}), the indicator variable ($Loss_{it}$) is excluded from the second stage model. In all four columns, we control for company size using the log of sales ($LnSale_{it}$), which means that there are no exclusion restrictions on any company size variable. Columns (3) and (4) are the ordinary probit models that do not control for selection bias (i.e., they do not include the $MILLS_{it}$ variables). Except for the exclusion of $MILLS_{it}$, the model specification in Column (3) is exactly the same as in Column (1), and Column (4) is the same as Column (2).

In Column (1), we obtain a significant negative coefficient for the Big_{it} variable, which corroborates the inferences of Columns (3) to (4) that do not control for selection bias. However, a different result is obtained in Column (2), in which the Big_{it} coefficient flips its sign, becoming positive and statistically insignificant. The $MILLS_{it}$ coefficient is also insignificant in Column (2), which suggests no selection bias; yet, the $MILLS_{it}$ coefficient in Column (1) is positive and highly significant. Again, the Big_{it} coefficients have similar magnitudes in the models that do not control for selection bias (−0.291 and −0.230), whereas they are widely different in the second stage

⁵ Multicollinearity is typically regarded as high (very high) when the variance-inflation-factors (VIFs) exceed 10 (20) (Belsley at al. 1980; Greene 2008).

⁶ In untabulated analyses, we also examine the relation between auditor choice and abnormal accruals using the same exclusion restrictions on company size that are used in Panels A and B of Table 4. We again find that the selection model yields results that are fragile and non-robust and that multicollinearity is high. In contrast, the OLS models have robust results and multicollinearity is low.

models that control for selection bias (-1.295 and 0.014). The VIFs are consistently high in the two selection models, whereas they are low in the ordinary probit models that do not control for selection bias. In the ordinary probit models, the Big_{it} coefficients are consistently negative and statistically significant and there are no multicollinearity problems. In short, Table 4, Panel C demonstrates that results can be fragile in the selection model when researchers impose exclusion restrictions on different measures of profitability. In contrast, the results are robust to different measures of profitability when the researcher does not attempt to control for selection bias.

Although $MILLS_{it}$ is insignificant in Column (2) of Panel C, dropping this control for selection bias dramatically changes the inferences for the treatment variable, Big_{it} . This is shown in Column (4), which has the same specification as Column (2) except there is no control for selection bias (i.e., the $MILLS_{it}$ variable is dropped). The Big_{it} coefficient is significantly negative in Column (4), whereas it is insignificantly positive in Column (2). Therefore, even when the $MILLS_{it}$ coefficient is insignificant—causing most researchers to conclude that there is no problem with selection bias—the decision to drop this control for selection bias dramatically changes the inferences for the treatment variable. Thus, high multicollinearity and the resulting lack of significance for the $MILLS_{it}$ coefficient greatly complicate the inferences for Big_{it} .

Next, we analyze the application of the selection model to the situation where D is a partitioning variable rather than a treatment variable. To do this, we estimate second stage models that have exactly the same specifications as in Table 4, but the models are estimated separately on the clients of Big 5 and non-Big 5 auditors. This analysis is different because Big_{it} is a treatment variable in Table 4, whereas the Big_{it} variable is used to partition the estimation samples in Table 5. Consistent with Table 4, we continue to find that the $MILLS_{it}$ results are unstable across the different model specifications in all three panels. For example, in Panel A the $MILLS_{it}$ coefficients for Big 5 clients are significantly positive in Columns (1) and (3) but significantly negative in Columns (2) and (4). Similarly, in Panel C the $MILLS_{it}$ coefficients for non-Big 5 clients are significantly positive in Column (1) but significantly negative in Column (2). Therefore, inferences as to the direction of the estimated selection bias are fragile. In addition, multicollinearity still poses a significant problem, particularly in the specifications that fail to impose exclusion restrictions. For example, two of the model specifications in Panel B have VIFs of 340.44 and 411.85.

In sum, the examples in Tables 4 and 5 illustrate that the results obtained from selection models are extremely sensitive when the researcher chooses no exclusion restrictions or chooses *ad hoc* restrictions with respect to company size and profitability. Importantly, the restrictions that we choose are typical of those found in accounting studies that attempt to control for endogeneity in auditor choice. In particular, some articles impose exclusion restrictions on profitability variables, some impose restrictions on company size variables, whereas other studies impose no exclusion restrictions at all. Moreover, most prior studies provide no clear justification or economic intuition for their chosen restrictions. The takeaway from our analysis is that the selection model choices we examine yield inferences that are too fragile to be trusted.

V. REPLICATING A RECENTLY PUBLISHED STUDY

The empirical examples of the previous section are motivated by research design choices commonly found in the literature. However, they are “homemade” in the sense that they do not directly replicate any of the studies in our survey. A skeptic may argue that studies published in top-tier journals are unlikely to contain such glaring problems given that they must pass the scrutiny of experienced editors and reviewers. Therefore, we replicate a published study and investigate whether its inferences are robust. Another motivation for this further analysis is to determine whether our critique is limited to the audit literature or if it applies more generally.

TABLE 5

Estimating the Sample Selection Model when Auditor Choice is a Partitioning Variable

Panel A: The Relation between Audit Fees and Auditor Size

The Second Stage Audit Fee Models for Big 5 Clients ($Big_{it} = 1$)				
	(1)	(2)	(3)	(4)
$MILLS_{it}$	0.327*** (7.73)	-0.369*** (-8.94)	1.206*** (19.51)	-0.103*** (-3.09)
Variance-Inflation-Factors (VIFs)				
$MILLS_{it}$	3.95	3.02	6.26	3.01
The Second Stage Audit Fee Models for Non-Big 5 Clients ($Big_{it} = 0$)				
	(1)	(2)	(3)	(4)
$MILLS_{it}$	-0.245*** (-4.40)	-0.930*** (-25.42)	-0.972*** (-12.09)	-0.517*** (-13.10)
Variance-Inflation-Factors (VIFs)				
$MILLS_{it}$	9.83	3.26	12.72	4.67

The model specifications in Columns (1) to (4) are the same as in Columns (1) to (4) of Panel A in Table 4.

Panel B: The Relation between the Cost of Equity and Auditor Size

The Second Stage Cost-of-Equity Models for Big 5 Clients ($Big_{it} = 1$)				
	(1)	(2)	(3)	(4)
$MILLS_{it}$	0.007** (2.07)	0.033*** (9.26)	-0.002 (-0.62)	-0.029*** (-7.70)
Variance-Inflation-Factors (VIFs)				
$MILLS_{it}$	3.94	3.24	4.68	3.17
The Second Stage Cost-of-Equity Models for Non-Big 5 Clients ($Big_{it} = 0$)				
	(1)	(2)	(3)	(4)
$MILLS_{it}$	-0.113*** (-3.55)	0.020*** (3.92)	-0.047 (-1.20)	-0.021*** (-4.10)
Variance-Inflation-Factors (VIFs)				
$MILLS_{it}$	340.44	7.37	411.85	7.02

The model specifications in Columns (1) to (4) are the same as in Columns (1) to (4) of Panel B in Table 4.

(continued on next page)

TABLE 5 (continued)

Panel C: The Relation between Going-Concern Audit Opinions and Auditor Size

The Second Stage Going-Concern Audit Opinion Models for Big 5 Clients ($Big_{it} = 1$)			
	(1)		(2)
$MILLS_{it}$	0.614*** (4.95)		-0.070 (-0.51)
Variance-Inflation-Factors (VIFs)			
$MILLS_{it}$	6.22		5.87
The Second Stage Going-Concern Audit Opinion Models for Non-Big 5 Clients ($Big_{it} = 0$)			
	(1)		(2)
$MILLS_{it}$	0.482*** (3.45)		-0.358*** (-2.78)
Variance-Inflation-Factors (VIFs)			
$MILLS_{it}$	11.74		10.47

The model specifications in Columns (1) and (2) are the same as in Columns (1) and (2) of Panel C in Table 4.

, * Significant at the 5 percent and 1 percent levels, respectively (two-tailed).

The models in this table have the same specifications as in Table 4, but they are estimated separately on the clients of Big 5 auditors and the clients of non-Big 5 auditors.

Variable Definitions:

$Big_{it} = 1$ if the company’s auditor is a Big 5 firm, and 0 otherwise; and

$MILLS_{it}$ = inverse Mills’ ratio calculated from the auditor-choice models in Table 3.

The study we use for this investigation is Jackson et al. (2009; hereafter, JLC).⁷ The point of this analysis is not to prove that the JLC study reached the wrong conclusion or to identify the correctly specified model for their particular research question. Rather, our objective is to show that their conclusions are fragile and inconclusive because the results can point in any direction with fairly minor changes to the set of exclusion restrictions.

JLC examine the impact of companies’ depreciation methods on capital investment decisions. They predict and report that companies using accelerated depreciation methods have higher levels of capital investments. Because a company’s choice of depreciation policy is endogenous, JLC first estimate a model that explains this decision. They then construct the inverse Mills’ ratio and include it as an independent variable in their second stage model of capital expenditure. Most of the independent variables from their first stage model are excluded from the second stage model of capital expenditure. As is typical of many studies in Table 2, JLC do not explain why they consider

⁷ Naturally, we cannot replicate all the studies that feature in our survey. Instead we select just one where our choice is motivated by the following considerations. First, we ignore the three studies that justify their chosen exclusion restrictions. Second, we ignore the 43 studies that estimate a selection model as part of a corroborative supplementary analysis because our objective is to determine whether the study’s primary inferences are fragile. Third, in the interest of transparency, we require that the study uses publicly available data from Compustat/CRSP in order that the reader can easily verify and replicate our findings. The vast majority of studies utilize one or more variables that are unavailable so this third screen leaves us with few candidates. Finally, we choose a study published in *JAR*, *JAE*, or *TAR* as these journals are generally regarded as the “top three.” These four screens leave us with a single study: Jackson et al. (2009).

their exclusion restrictions to be valid, nor do they report diagnostic tests for multicollinearity or sensitivity analyses for robustness.

We first verify that that our sample descriptive statistics are similar to JLC and that we obtain similar results when estimating their first and second stage models. JLC do not report diagnostic tests for multicollinearity, but we do. The VIFs for the depreciation treatment variable ($CHOICE_{it}$) and the inverse Mills ratio ($MILLS_{it}$) are 19.56 and 23.34 indicating that multicollinearity is high in their model specifications. (All untabulated results are available from the authors.)

Our main focus is on whether the results in JLC's second stage model are robust to their exclusion restrictions. One of JLC's exclusion restrictions involves a variable titled $LABOR_{it}$, which is measured as 1 minus the ratio of gross property, plant, and equipment (PPE) to average total assets. The variable $LABOR$ captures the firm's capital intensity, with a larger value of $LABOR$ denoting firms with lower capital intensity (i.e., less PPE). JLC include $LABOR_{it}$ in the first stage model of depreciation choice and exclude it from their second stage model of capital expenditure. However, $LABOR_{it}$ is mechanically and contemporaneously correlated with the second stage dependent variable ($CAPX_{it}$) because capital expenditures in year t result in higher values of gross PPE and therefore lower values of $LABOR_{it}$. This is problematic and occurs because both $LABOR_{it}$ and $CAPX_{it}$ variables are constructed using the level of PPE in year t . In other words, JLC's $MILLS_{it}$ variable embeds $LABOR_{it}$ from the first stage model, and thereby induces a mechanical contemporaneous correlation between $MILLS_{it}$ and $CAPX_{it}$ in the second stage.

As emphasized in the JLC study, there are compelling theoretical and empirical reasons to believe that some measure of $LABOR$ (a proxy for capital intensity) should be included in the first stage model of depreciation. The question then is how to do this without inducing the mechanical contemporaneous correlation problem. We believe the best answer is to replace $LABOR_{it}$ with $LABOR_{it-2}$ in the first stage model because there is no mechanical contemporaneous relation between $LABOR_{it-2}$ and $CAPX_{it}$, given that these variables are measured in different time periods. In addition, it can be argued that $LABOR_{it-2}$ should also be included in the second stage model because companies that relied relatively more on physical capital as an input to production in the past would incur larger capital expenditures as their old assets are replaced. Therefore, companies that have historically high levels of PPE would make larger capital investments in year t , implying that $LABOR_{it-2}$ is an important determinant of capital expenditure.

The results of this sensitivity analysis are reported in Table 6. Column (2) reports our modified version of the second stage model that includes $LABOR_{it-2}$ ($LABOR_{it-2}$ is also included in the first stage model). The OLS specification in Column (3) is the same as Column (2) except that it does not include the $MILLS_{it}$ variable. For the purposes of comparison, Column (1) presents the results reported by JLC. In Column (2), the $CHOICE_{it}$ and $MILLS_{it}$ coefficients in our model are statistically insignificant, which contrasts sharply with the highly significant results reported by JLC. The coefficient for $CHOICE_{it}$ is also insignificant in the OLS model, which is consistent with our selection model in Column (2), but is inconsistent with JLC's selection model. The VIFs for $CHOICE_{it}$ and $MILLS_{it}$ are very high at 19.56 and 23.34, whereas the VIF for $CHOICE_{it}$ is only 1.29 in the OLS model. Thus, multicollinearity is high in the selection model and very low in the OLS model.⁸

Another potential solution to the mechanical correlation problem would be to remove $LABOR$ entirely from the first and second stage models. This solution is likely inferior to including

⁸ In an untabulated test, we relax JLC's exclusion restriction on $LABOR_{it}$ by including this variable in both the first and second stage models. In this specification, the $CHOICE_{it}$ coefficient becomes significantly negative, whereas it is significantly positive in JLC. The results also change dramatically for the $MILLS_{it}$ variable. The $MILLS_{it}$ coefficient is positive and insignificant when $LABOR_{it}$ is included in the second stage model, whereas the $MILLS_{it}$ coefficient is significantly negative when JLC exclude $LABOR_{it}$ from the second stage model.

TABLE 6
Sensitivity Analysis for the Second Stage Model of JLC's Study

	JLC's Results		Our Sensitivity Analysis	
	(1)	(2)	(2)	(3)
<i>CHOICE_{it}</i>	0.038*** (12.81)	0.001 (0.23)	0.001 (0.23)	-0.000 (-0.59)
<i>MILLS_{it}</i>	-0.021*** (-13.09)	-0.001 (-0.33)	-0.001 (-0.33)	-0.001 (-0.33)
<i>LABOR_{it-2}</i>		-0.034*** (-18.52)	-0.034*** (-18.52)	-0.034*** (-20.68)
<i>CAPX_{it-1}</i>	0.536*** (70.31)	0.401*** (30.29)	0.401*** (30.29)	0.401*** (30.30)
<i>MB_{it-1}</i>	0.002*** (12.80)	0.002*** (12.39)	0.002*** (12.39)	0.002*** (12.41)
<i>LEVTA_{it-1}</i>	-0.017*** (-13.27)	-0.014*** (-10.78)	-0.014*** (-10.78)	-0.014*** (-10.79)
<i>CASH_{it-1}</i>	-0.007*** (-6.33)	-0.003** (-2.13)	-0.003** (-2.13)	-0.003** (-2.14)
Δ <i>CASH_{it}</i>	-0.029*** (-15.60)	-0.023*** (-11.56)	-0.023*** (-11.56)	-0.023*** (-11.56)
<i>AGE_{it-1}</i>	-0.001*** (-6.83)	-0.002*** (-8.52)	-0.002*** (-8.52)	-0.002*** (-8.56)
<i>SIZE_{it-1}</i>	0.000 (0.19)	0.000 (1.06)	0.000 (1.06)	0.000 (1.06)
<i>RET_{it-1}</i>	0.003*** (15.60)	0.002*** (10.66)	0.002*** (10.66)	0.002*** (10.67)
<i>CFO_{it}</i>	0.022*** (14.00)	0.015*** (9.14)	0.015*** (9.14)	0.015*** (9.20)
Δ <i>SALE_{it}</i>	0.018*** (26.17)	0.016*** (22.75)	0.016*** (22.75)	0.016*** (22.80)
Intercept	0.015*** (4.25)	0.047*** (10.22)	0.047*** (10.22)	0.048*** (10.54)
Observations	61,096	50,656	50,656	50,656
Pseudo R ² (%)	49.58	47.15%	47.15%	47.15%
Variance-Inflation-Factors (VIFs)				
<i>CHOICE_{it}</i>	Not reported	28.39	28.39	1.29
<i>MILLS_{it}</i>	Not reported	23.59	23.59	

** , *** Significant at the 5 percent and 1 percent levels, respectively (two-tailed).

The standard errors are adjusted for time-series dependence by clustering on each company (Petersen 2009).

In our sensitivity analysis, *LABOR_{it-2}* is included in the first stage and second stage models instead of *LABOR_{it}*. *LABOR_{it}* is omitted from the first stage and second stage models in order to avoid a mechanical contemporaneous correlation between *MILLS_{it}* and *CAPX_{it}*. Except for this, the models are the same as in JLC (t-statistics are reported in parentheses).

Variable Definitions:

CHOICE_{it} = 1 for companies that use the accelerated depreciation method for all or some of their assets, and 0 for companies that use the straight-line depreciation method;

LABOR_{it} = labor intensity measured as 1 minus the ratio of gross property, plant, and equipment to *ADJ_TA_{it}*;

CAPX_{it} = capital investments scaled by *ADJ_TA_{it}*;

MB_{it} = market value of common equity plus book value of assets minus book value of equity, with the resulting amount scaled by *ADJ_TA_{it}*;

LEVTA_{it} = the book value of short-term debt plus book value of long-term debt, with this sum scaled by *ADJ_TA_{it}*;

(continued on next page)

TABLE 6 (continued)

$CASH_{it}$ = cash and short-term investments scaled by ADJ_TA_{it} ;

$\Delta CASH_{it}$ = year-to-year change in $CASH_{it}$;

AGE_{it} = log of the number of years the company has been listed on CRSP;

$SIZE_{it}$ = log of ADJ_TA_{it} ;

RET_{it} = stock returns;

CFO_{it} = cash flows from operations scaled by ADJ_TA_{it} ;

$\Delta SALE_{it}$ = year-to-year change in net sales scaled by ADJ_TA_{it} ;

ADJ_TA_{it} = average of beginning of year and end of year adjusted total assets, where adjusted total assets are equal to total assets plus accumulated depreciation; and

i and t = company and year subscripts, respectively.

$LABOR_{it-2}$ because there are compelling economic reasons to expect that companies with higher levels of PPE would need to invest more on maintaining their capital stock. Further, $LABOR_{it-2}$ is highly significant in the first and second stage models, as one would expect, so omitting this variable could create an omitted variable problem. Nevertheless, if we exclude both $LABOR_{it-2}$ (and $LABOR_{it}$) from the first and second stage models, and keep all other exclusion restrictions exactly the same as in JLC's study, we revert back to their finding of a significant positive coefficient on $CHOICE_{it}$ (coefficient = 0.009) and a significant negative coefficient on $MILLS_{it}$ (coefficient = -0.005). Notwithstanding their statistical significance, these coefficients are much smaller in magnitude than in JLC where the $CHOICE_{it}$ coefficient is 0.038 and the $MILLS_{it}$ coefficient is -0.021. In fact, we have been unable to find any model specifications that give stronger results than those reported by JLC.

In addition, even when $LABOR_{it}$ and $LABOR_{it-2}$ are excluded from the first and second stage models, we find that the significant positive coefficient on $CHOICE_{it}$ and the significant negative coefficient on $MILLS_{it}$ are not robust to some of the other exclusion restrictions that JLC impose in the second stage. In untabulated tests, we find that some specifications yield results that are consistent with but weaker than the JLC study, whereas in other specifications, their inferences disappear completely. For example, in untabulated analyses, we investigate the consequences of relaxing JLC's exclusion restrictions over sales ($SALE_{it}$), an interaction between sales and an indicator for the oil and gas industry ($OG_{it} \times SALE_{it}$), and leverage (LEV_{it}), while retaining all their other restrictions.⁹

Once again, we find that results for the treatment variable ($CHOICE_{it}$) and the test of selection bias ($MILLS_{it}$) are often inconsistent with the JLC study. In one specification, the $CHOICE_{it}$ coefficient is significantly negative ($p < 0.10$) and the $MILLS_{it}$ coefficient is significantly positive ($p < 0.05$), which are opposite to the signs reported by JLC. In three other specifications, the $CHOICE_{it}$ and $MILLS_{it}$ coefficients are statistically insignificant. Thus, the fragility of the selection model with respect to the exclusion of $LABOR_{it}$ or $LABOR_{it-2}$ extends to the exclusion of other variables that can be justifiably included in the second stage model. While we cannot conclude that JLC's inferences are necessarily incorrect, the JLC study does not justify its chosen exclusion

⁹ The exclusion restrictions that JLC impose on $SALE_{it}$ and $OG_{it} \times SALE_{it}$ are questionable because companies with high sales have more reason to invest in new capital, and such investments are likely to be particularly large in the capital-intensive oil and gas industry. Their restriction over LEV_{it} is also questionable because highly leveraged companies are likely to have difficulty raising external finance, implying that such companies would have lower capital expenditures. Consistent with these arguments, we find significant positive coefficients for $SALE_{it}$ and $OG_{it} \times SALE_{it}$ and a significant negative coefficient for LEV_{it} in the second stage models of capital expenditure.

restrictions and their results are sensitive to alternative and reasonable restrictions, implying that their inferences are not robust.

VI. SUGGESTIONS FOR BETTER IMPLEMENTATION OF SELECTION MODELS

We believe that implementations of the selection model in the accounting literature can be greatly improved. We offer four practical suggestions for accounting researchers, and for editors and reviewers in evaluating the use of selection models. First, it is inadvisable to estimate selection models without exclusion restrictions (i.e., some of the independent variables from the first stage model should be excluded from the second stage model). In the absence of exclusion restrictions, the results for the inverse Mills' ratio depend entirely on its nonlinearity. This is a problem because theory rarely provides information about the correct functional form. We find it difficult to envisage an accounting setting in which the empirical researcher can be sure that the nonlinearity of the inverse Mills' ratio is capturing the effects of selection bias rather than misspecification of the functional form. And, because multicollinearity can arise even when exclusion restrictions are imposed, we recommend that researchers report diagnostic tests for multicollinearity.

Second, some accounting studies do not report which independent variables are used in the first stage models. This lack of disclosure means that the reader cannot identify the set of exclusion restrictions or evaluate their statistical power. Because exclusion restrictions are fundamental to the resulting inferences, our second recommendation is that studies need to explicitly report the results from their first stage choice models and clearly identify their chosen Z variables.

Third, studies should justify why the Z variables in the first stage model can be validly excluded from the second stage model. Accounting researchers are accustomed to explaining why certain independent variables are *included* in a model, but they often fail to explain why one or more variables from the first stage choice model can be validly *excluded* from the second stage outcome model. It is not enough to rely on past studies to justify the independent variables that are included in the first and second stage models. More important is the need for researchers to justify and explain which of the independent variables in the first stage model can be validly *excluded* from the second stage model.

Fourth, because the selection models are often fragile, it is essential to report sensitivity analyses in order to establish the robustness of inferences. It is surprising that this practice is uncommon, given that this is the norm in most other types of empirical accounting research. But given its many implementation challenges, [Bushway et al. \(2007, 151\)](#) suggest "Thoughtful consideration is therefore needed before employing this common but overused technique."

In our review of the accounting literature, there are some positive examples of selection models. For example, [Bushee et al. \(2003\)](#) stands out in terms of justifying the variables in the first and second stage models, reporting sensitivity tests to alternative model specifications, and reporting diagnostic tests for multicollinearity. Moreover, [Bushee et al. \(2003\)](#) clearly identify their exclusion restrictions and they state that they do not have good reason to expect that those Z variables would directly affect the dependent variable in the second stage model. Likewise, [Feng et al. \(2009\)](#) emphasize the importance of having at least one exclusion restriction, report diagnostic tests for multicollinearity, and investigate whether their results are sensitive to alternative model specifications.

Outside of accounting, [Angrist \(1990\)](#) provides one of the best known choices of an exclusion restriction. [Angrist \(1990\)](#) examines how military service (D) affects the civilian earnings of veterans after they are discharged (Y). This involves a selection issue because the decision to serve in the military (D) is endogenous and the researcher cannot observe what soldiers would have earned if they had chosen a different profession. [Angrist \(1990\)](#) tackles this selection problem using data from the Vietnam era, when priority for military conscription was allocated through a draft

lottery. At first glance, the lottery seems to be an ideal Z variable because: (1) it is random, (2) it is a powerful predictor of military service (D), and (3) it is hard to argue that the lottery would have a direct effect on civilian earnings (Y). However, even the lottery is not completely immune from criticism as a Z variable. Angrist (1990) points out that that, during the Vietnam era, some students went to college in order to avoid the draft. Draft-avoidance behavior (i.e., going to college) is likely correlated with the assigned lottery number (Z) because it was common knowledge that people with lower draft numbers were more likely to be drafted. Moreover, the decision to go to college and avoid the draft is likely endogenous to expected future earnings (Y). Thus, the lottery (Z) could very well be correlated with the error term in the second stage Y model, in which case it would not be exogenous (Heckman 1997).

What should a researcher do if defensible exclusion restrictions are not available? In our view, the appropriate recommendation depends on one of two situations. The first case is where the researcher's primary inferences are obtained without controlling for selection bias but there is a concern that those results may be biased. In this situation, our empirical examples illustrate that it is relatively easy to find an *ad hoc* specification of the selection model that corroborates the OLS findings. However, such corroborative findings provide little assurance that selection bias exists or has truly been eliminated, as they are easily overturned using alternative and equally plausible specifications of the selection model. In this situation, the researcher should explain that a potential endogeneity problem could affect the study's inferences, but that it is very difficult to implement a credible selection model. In our view, this is better science than making unreliable claims of having controlled for selection bias. In turn, we believe that reviewers and editors should be more hesitant about pushing authors to use the selection model to corroborate the results from OLS. A helpful reviewer would try to think of a suitable exclusion restriction or, where none exists, the reviewer should admit that s/he is unable to come up with a good one. In contrast, our experience is that reviewers ask authors to estimate a selection model without advising them as to the appropriate exclusion restrictions. Authors are understandably reluctant to go against a reviewer's suggestion even when they have good reason. Thus, even when authors are unable to find justifiable exclusion restrictions, they feel compelled to find a specification of the selection model that corroborates their OLS findings. Our point is that this is not good science.

The second situation is where the researcher obtains results using the selection model with arbitrary exclusion restrictions or no restrictions at all and the results of the selection model are different from those obtained using OLS. This situation is arguably even worse than the first, because then the study's inferences depend entirely on an *ad hoc* specification of the selection model with results that are inconsistent with those obtained using OLS. In such situations, authors should not assume that the selection model is better than OLS. Instead, authors should disclose that the results point in different directions and they should evaluate the relative fragility of inferences obtained using the selection and OLS models.

Our empirical examples illustrate that OLS is typically more robust because it does not require exclusion restrictions and multicollinearity is generally low because the inverse Mills' ratio is excluded. However, we emphasize that dropping the inverse Mills' ratio is not always preferable to using a selection model. Although OLS is typically more robust, it can still yield incorrect inferences when selection bias is a significant concern. Nevertheless, robustness is an important criterion that researchers should take into account when evaluating their findings, a point that is already well recognized in the accounting literature as well as by econometricians. We echo the comment by Leamer (1983, 38) that:

An inference is not believable if it is fragile, if it can be reversed by minor changes in assumptions. As consumers of research, we correctly reserve judgment on an inference until it stands up to a study of fragility, usually by other researchers advocating opposite

opinions. It is, however, much more efficient for individual researchers to perform their own sensitivity analyses, and we ought to be demanding much more complete and more honest reporting of the fragility of claimed inferences.

VII. CONCLUSIONS

Our review of the accounting literature indicates that some studies have implemented the selection model in a questionable manner. Accounting researchers often impose *ad hoc* exclusion restrictions or no exclusion restrictions whatsoever. Using empirical examples and a replication of a published study, we demonstrate that such practices can yield results that are too fragile to be considered reliable. In our empirical examples, a researcher could obtain quite literally *any* outcome by making relatively minor and apparently innocuous changes to the set of exclusionary variables, including choosing a null set. One set of exclusion restrictions would lead the researcher to conclude that selection bias is a significant problem, while an alternative set involving rather minor changes would give the opposite conclusion. Thus, claims about the existence and direction of selection bias can be sensitive to the researcher's set of exclusion restrictions.

Our examples also illustrate that the selection model is vulnerable to high levels of multicollinearity, which can exacerbate the bias that arises when a model is misspecified (Thursby 1988). Moreover, the potential for misspecification is high in the selection model because inferences about the existence and direction of selection bias depend entirely on the researcher's assumptions about the appropriate functional form and exclusion restrictions. In addition, high multicollinearity means that the statistical insignificance of the inverse Mills' ratio is not a reliable guide as to the absence of selection bias. Even when the inverse Mills' ratio is statistically insignificant, inferences from the selection model can be different from those obtained without the inverse Mills' ratio. In this situation, the selection model indicates that it is legitimate to omit the inverse Mills' ratio, and yet, omitting the inverse Mills' ratio gives different inferences for the treatment variable because multicollinearity is then much lower.

In short, researchers are faced with the following trade-off. On the one hand, selection models can be fragile and suffer from multicollinearity problems, which hinder their reliability. On the other hand, the selection model potentially provides more reliable inferences by controlling for endogeneity bias if the researcher can find good exclusion restrictions, and if the models are found to be robust to minor specification changes. The importance of these advantages and disadvantages depends on the specific empirical setting, so it would be inappropriate for us to make a general statement about when the selection model should be used. Instead, researchers need to critically appraise the quality of their exclusion restrictions and assess whether there are problems of fragility and multicollinearity in their specific empirical setting that might limit the effectiveness of selection models relative to OLS.

Another way to control for unobservable factors that are correlated with the endogenous regressor (D) is to use panel data. Though it may be true that many unobservable factors impact the choice of D , as long as those unobservable characteristics remain constant during the period of study, they can be controlled for using a fixed effects research design. In this case, panel data tests that control for unobserved differences between the treatment group ($D = 1$) and the control group ($D = 0$) will eliminate the potential bias caused by endogeneity as long as the unobserved source of the endogeneity is time-invariant (e.g., Baltagi 1995; Meyer 1995; Bertrand et al. 2004). The advantages of such a difference-in-differences research design are well recognized by accounting researchers (e.g., Altamuro et al. 2005; Desai et al. 2006; Hail and Leuz 2009; Hanlon et al. 2008). As a caveat, however, we note that the time-invariance of unobservables is a strong assumption that cannot be empirically validated. Moreover, the standard errors in such panel data tests need to be

corrected for serial correlation because otherwise there is a danger of over-rejecting the null hypothesis that D has no effect on Y (Bertrand et al. 2004).¹⁰

Finally, we note that there is a recent trend in the accounting literature to use samples that are matched based on their propensity scores (e.g., Armstrong et al. 2010; Lawrence et al. 2011). An advantage of propensity score matching (PSM) is that there is no *MILLS* variable and so the researcher is not required to find valid Z variables (Heckman et al. 1997; Heckman and Navarro-Lozano 2004). However, such matching has two important limitations. First, selection is assumed to occur only on observable characteristics. That is, the error term in the first stage model is correlated with the independent variables in the second stage (i.e., u is correlated with X and/or Z), but there is no selection on unobservables (i.e., u and v are uncorrelated). In contrast, the purpose of the selection model is to control for endogeneity that arises from *unobservables* (i.e., the correlation between u and v). Therefore, propensity score matching should not be viewed as a replacement for the selection model (Tucker 2010).

A second limitation arises if the treatment variable affects the company's matching attributes. For example, suppose that a company's choice of auditor affects its subsequent ability to raise external capital. This would mean that companies with higher quality auditors would grow faster. Suppose also that the company's characteristics at the time the auditor is first chosen cannot be observed. Instead, we match at some stacked calendar time where some companies have been using the same auditor for 20 years and others for not very long. Then, if we matched on company size, we would be throwing out the companies that have become large because they have benefited from high-quality audits. Such companies do not look like suitable "matches," insofar as they are much larger than the companies in the control group that have low-quality auditors. In this situation, propensity matching could bias toward a non-result because the treatment variable (auditor choice) affects the company's matching attributes (e.g., its size). It is beyond the scope of this study to provide a more thorough assessment of the advantages and disadvantages of propensity score matching in accounting applications, so we leave this important issue to future research.

REFERENCES

- Ahmed, S., E. Kilic, and G. Lobo. 2006. Does recognition versus disclosure matter? Evidence from value-relevance of banks' recognized and disclosed derivative financial instruments. *The Accounting Review* 81 (3): 567–588.
- Altamuro, J., A. Beatty, and J. Weber. 2005. The effects of accelerated revenue recognition on earnings management and earnings informativeness: Evidence from SEC Staff Accounting Bulletin No. 101. *The Accounting Review* 80 (2): 373–401.
- Angrist, J. 1990. Lifetime earnings and the Vietnam era draft lottery: Evidence from Social Security Administrative records. *American Economic Review* 80 (3): 313–316.
- Armstrong, C., A. Jagolinzer, and D. 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 contracts. *Journal of Accounting and Economics* 30 (3): 245–278.

¹⁰ It is sometimes possible to address potential endogeneity threats in other ways. For example, in their study of insider trading activity and market reaction to accounting restatements, Badertscher et al. (2011) conduct the following sensitivity test. They reason that if the market reaction is truly conditioned on insider trading, then their results would only be found in those firms where the insider activity is announced before the restatement, and not when insider trades occur before restatements but are not announced until after the restatements. The fact that they see an effect only for the insider activity announced before the restatement adds credibility to their primary results and lessens the concern over selection bias. We thank Steve Kachelmeier for bringing this study to our attention.

- Atiase, R., H. Li, S. Supattarakul, and S. Tse. 2005. Market reaction to multiple contemporaneous earnings signals: Earnings announcements and future earnings guidance. *Review of Accounting Studies* 40 (4): 497–525.
- Baber, W., W. Chen, and S.-H. Kang. 2006. Stock price reaction to evidence of earnings management: Implications for supplementary financial disclosure. *Review of Accounting Studies* 11 (1): 5–19.
- Baber, W., and A. Gore. 2008. Consequences of GAAP disclosure regulation: Evidence from municipal debt issues. *The Accounting Review* 83 (3): 565–591.
- Badertscher, B., P. Hribar, and N. T. Jenkins. 2011. Informed Trading and the Market Reaction to Accounting Restatements. *The Accounting Review* 86 (5): 1491–1518.
- Baginski, S., J. Hassell, and M. Kimbrough. 2004. Why do managers explain their earnings forecasts? *Journal of Accounting Research* 42 (1): 1–29.
- Ball, R., and L. Shivakumar. 2005. Earnings quality in UK private firms: Comparative loss recognition timeliness. *Journal of Accounting and Economics* 39 (1): 83–128.
- Baltagi, B. H. 1995. *Econometric Analysis of Panel Data*. Chichester, U.K.: Wiley.
- Barton, J. 2001. Does the use of financial derivatives affect earnings decisions? *The Accounting Review* 76 (1): 1–26.
- Beatty, A., K. Ramesh, and J. Weber. 2002. The importance of accounting changes in debt contracts: The cost of flexibility in covenant calculations. *Journal of Accounting and Economics* 33 (2): 205–227.
- Behn, B., J.-H. Choi, and T. Kang. 2008. Audit quality and properties of analyst earnings forecasts. *The Accounting Review* 83 (2): 327–349.
- Belsley, D., E. Kuh, and E. R. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York, NY: Wiley & Sons.
- Bens, D., and R. Johnston. 2009. Accounting discretion: Use or abuse? An analysis of restructuring charges surrounding regulator action. *Contemporary Accounting Research* 26 (3): 673–99.
- Bertrand, M., E. Duflo, and S. Mullainathan. 2004. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics* 119 (1): 249–275.
- Blaconiere, W., M. Johnson, and M. Lewis. 2008. The role of tax regulation and compensation contracts in the decision to voluntarily expense employee stock options. *Journal of Accounting and Economics* 46 (1): 101–111.
- Blouin, J., and L. Krull. 2009. Bringing it home: A study of the incentives surrounding the repatriation of foreign earnings. *Journal of Accounting Research* 47 (4): 1027–1059.
- Bushee, B., D. Matsumoto, and G. Miller. 2003. Open versus closed conference calls: The determinants and effects of broadening access to disclosure. *Journal of Accounting and Economics* 34 (1-3): 149–180.
- Bushway, S., B. Johnson, and L. Slocum. 2007. Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology. *Journal of Quantitative Criminology* 23 (2): 151–178.
- Butler, M., A. Kraft, and I. Weiss. 2007. The effect of reporting frequency on the timeliness of earnings: The cases of voluntary and mandatory interim reports. *Journal of Accounting and Economics* 43 (2-3): 181–217.
- Carter, M. E., C. D. Ittner, and S. Zechman. 2009. Explicit relative performance evaluation in performance-vested equity grants. *Review of Accounting Studies* 14 (2-3): 269–306.
- Chaney, P., D. Jeter, and L. Shivakumar. 2004. Self-selection of auditors and audit pricing in private firms. *The Accounting Review* 79 (1): 51–72.
- Chang, X., S. Dasgupta, and G. Hilary. 2009. The effect of auditor quality on financing decisions. *The Accounting Review* 84 (4): 1085–1117.
- Cheng, Q., and D. Farber. 2008. Earnings restatements, changes in CEO compensation, and firm performance. *The Accounting Review* 83 (5): 1217–1250.
- Choi, J. H., J.-B. Kim, X. Liu, and D. Simunic. 2008. Audit pricing, legal liability regimes, and Big 4 premiums: Theory and cross-country evidence. *Contemporary Accounting Research* 25 (1): 55–99.
- Choi, J. H., J.-B. Kim, X. Liu, and D. D. Simunic. 2009. Cross-listing audit fee premiums: Theory and evidence. *The Accounting Review* 84 (5): 1429–1463.
- Christensen, H., E. Lee, and M. Walker. 2009. Do IFRS reconciliations convey information? The effect of debt contracting. *Journal of Accounting Research* 47 (5): 1167–1199.

- Chung, H., and J. Wynn. 2008. Managerial legal liability coverage and earnings conservatism. *Journal of Accounting and Economics* 46 (1): 135–153.
- Claus, J., and J. Thomas. 2001. Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets. *The Journal of Finance* 56 (5): 1629–1666.
- Cowen, A., B. Groyberg, and P. Healy. 2006. Which types of analyst firms are more optimistic? *Journal of Accounting and Economics* 41 (1-2): 119–146.
- Collins, D., O. Z. Li, and H. Xie. 2009. What drives the increased informativeness of earnings announcements over time? *Review of Accounting Studies* 13 (1): 1–30.
- Davila, A., and F. Penalva. 2006. Governance structure and the weighting of performance measures in CEO compensation. *Review of Accounting Studies* 11 (4): 463–493.
- DeFond, M., and M. Hung. 2007. Investor protection and analysts' cash flow forecasts around the world. *Review of Accounting Studies* 12 (2-3): 377–419.
- Desai, H., C. Hogan, and M. Wilkins. 2006. The reputational penalty for aggressive accounting: Earnings restatements and management turnover. *The Accounting Review* 81 (1): 83–112.
- Doyle, J., W. Ge, and S. McVay. 2007. Accruals quality and internal control over financial reporting. *The Accounting Review* 82 (5): 1141–1170.
- Easton, P. 2004. PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital. *The Accounting Review* 79 (1): 73–95.
- Eberhart, A., W. Maxwell, and A. Siddique. 2008. A reexamination of the tradeoff between the future benefit and riskiness of R&D increases. *Journal of Accounting Research* 46 (1): 27–52.
- Engel, E., R. Hayes, and X. Wang. 2007. The Sarbanes-Oxley Act and firms' going-private decisions. *Journal of Accounting and Economics* 44 (1-2): 116–145.
- Fee, E., and C. Hadlock. 2004. Management turnover across the corporate hierarchy. *Journal of Accounting and Economics* 37 (1): 3–38.
- Feng, M., C. Li, and S. McVay. 2009. Internal control and management guidance. *Journal of Accounting and Economics* 48 (2-3): 190–209.
- Files, R., E. Swanson, and S. Tse. 2009. Stealth disclosure of accounting restatements. *The Accounting Review* 84 (5): 1495–1520.
- Fortin, S., and J. Pittman. 2007. The role of auditor choice in debt pricing in private firms. *Contemporary Accounting Research* 24 (3): 859–96.
- Frankel, R., C. Seethamraju, and T. Zach. 2008. GAAP goodwill and debt contracting efficiency: Evidence from net-worth covenants. *Review of Accounting Studies* 13 (1): 87–118.
- Garfinkel, J., and J. Sokobin. 2006. Volume, opinion divergence, and returns: A study of post-earnings announcement drift. *Journal of Accounting Research* 44 (1): 85–112.
- Gebhardt, W., C. Lee, and B. Swaminathan. 2001. Toward an implied cost of capital. *Journal of Accounting Research* 39 (1): 135–176.
- Gode, D., and P. Mohanram. 2003. Inferring cost of capital using the Ohlson-Juettner model. *Review of Accounting Studies* 8 (4): 399–431.
- Gong, G., Y. Li, and H. Xie. 2009. The association between management earnings forecast errors and accruals. *The Accounting Review* 84 (2): 497–530.
- Greene, W. 2008. *Econometric Analysis*. Upper Saddle River, NJ: Pearson/Prentice Hall.
- Guedhami, O., and J. Pittman. 2006. Ownership concentration in privatized firms: The role of disclosure standards, auditor choice, and auditing infrastructure. *Journal of Accounting Research* 44 (5): 889–929.
- Hail, L., and C. Leuz. 2009. Cost of capital effects and changes in growth expectations around U.S. cross-listings. *Journal of Financial Economics* 93 (3): 428–454.
- Hanlon, M., E. Mayhew, and T. Shevlin. 2008. An unintended consequence of book-tax conformity: A loss of earnings informativeness. *Journal of Accounting and Economics* 46 (2-3): 294–311.
- Haw, I., B. Hu, L. Hwang, and W. Wu. 2004. Ultimate ownership, income management, and legal and extra-legal institutions. *Journal of Accounting Research* 42 (2): 23–462.
- Heckman, J. 1979. The sample selection bias as a specification error. *Econometrica* 47 (1): 153–62.

- Heckman, J. 1997. Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *The Journal of Human Resources* 32 (3): 441–462.
- Heckman, J., H. Ichimura, and P. Todd. 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economics and Statistics* 64 (4): 605–654.
- Heckman, J., and S. Navarro-Lozano. 2004. Using matching, instrumental variables, and control functions to estimate economic choice models. *The Review of Economics and Statistics* 86 (1): 30–57.
- Hoitash, U., R. Hoitash, and J. C. Bedard. 2009. Corporate governance and internal control over financial reporting: A comparison of regulatory regimes. *The Accounting Review* 84 (3): 839–867.
- Hui, K. W., S. Matsunaga, and D. Morse. 2009. The impact of conservatism on management earnings forecasts. *Journal of Accounting and Economics* 47 (3): 192–207.
- Jackson, S., X. Liu, and M. Cecchini. 2009. Economic consequences of firms' depreciation method choice: Evidence from capital investments. *Journal of Accounting and Economics* 48 (1): 54–68.
- Jiang, X. 2008. Beating earnings benchmarks and the cost of debt. *The Accounting Review* 83 (2): 377–416.
- Johnstone, K. M., J. C. Bedard, and M. L. Ettredge. 2004. The effect of competitive bidding on engagement planning and pricing. *Contemporary Accounting Research* 21 (1): 25–53.
- Katz, S. 2009. Earnings quality and ownership structure: The role of private equity sponsors. *The Accounting Review* 84 (3): 623–658.
- Ke, B., and Y. Yu. 2006. The effect of issuing biased earnings forecasts on analysts' access to management and survival. *Journal of Accounting Research* 44 (5): 965–999.
- Khurana, I., and 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.
- Kim, J.-B., R. Chung, and M. Firth. 2003. Auditor conservatism, asymmetric monitoring, and earnings management. *Contemporary Accounting Research* 20 (2): 323–359.
- Kim, J.-B., and C. H. Yi. 2006. Ownership structure, business group affiliation, listing status, and earnings management: Evidence from Korea. *Contemporary Accounting Research* 23 (2): 427–464.
- Kini, M., S. Mian, M. Rebello, and A. Venkateswaran. 2009. On the structure of analyst research portfolios. *Journal of Accounting Research* 47 (4): 867–909.
- Lang, M., K. Lins, and D. Miller. 2003. ADRs, analysts, and accuracy: Does cross listing in the United States improve a firm's information environment and increase market value? *Journal of Accounting Research* 41 (2): 317–345.
- Lara, G., G. Osma, and F. Penalva. 2009. Accounting conservatism and corporate governance. *Review of Accounting Studies* 14 (1): 161–201.
- Larcker, D., and T. Rusticus. 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics* 49 (3): 186–205.
- Lawrence, A., M. Minutti-Meza, and P. Zhang. 2011. Can Big 4 versus non-Big 4 differences in audit quality proxies be attributed to client characteristics? *The Accounting Review* 86 (1): 259–286.
- Leamer, E. 1983. Let's take the con out of econometrics. *American Economic Review* 73 (1): 31–43.
- Leuz, C. 2003. IAS versus US GAAP: Information asymmetry-based evidence from Germany's New Market. *Journal of Accounting Research* 41 (3): 445–472.
- Leuz, C., and R. Verrecchia. 2000. The economic consequences of increased disclosure. *Journal of Accounting Research* 38 (Supplement): 91–124.
- Li, K., and N. Prabhala. 2007. Self-selection models in corporate finance. In *Handbook of Corporate Finance: Empirical Corporate Finance*, edited by B. E. Eckso. Amsterdam, The Netherlands: North-Holland.
- Li, C. 2009. Does client importance affect auditor independence at the office level? Empirical evidence from going-concern opinions. *Contemporary Accounting Research* 26 (1): 201–30.
- Little, R. 1985. A note about models for selectivity bias. *Econometrica* 53 (6): 1469–1474.
- Little, R., and D. Rubin. 1987. *Statistical Analysis with Missing Data*. New York, NY: John Wiley & Sons.
- Louis, H. 2005. Acquirers' abnormal returns and the non-Big 4 auditor clientele effect. *Journal of Accounting and Economics* 40 (1-3): 75–99.
- Louis, H., D. Robinson, and A. Sbaraglia. 2008. An integrated analysis of the association between accrual disclosure and the abnormal accrual anomaly. *Review of Accounting Studies* 13: 23–54.

- Maddala, G. 1991. A perspective on the use of limited-dependent and qualitative variables models in accounting research. *The Accounting Review* 66 (4): 788–807.
- Manning, W., G. Duan, and W. Roger. 1987. Monte Carlo evidence on the choice between sample selection and two-part models. *Journal of Econometrics* 35 (1): 59–82.
- Mansi, S., W. Maxwell, and D. Miller. 2004. Does auditor quality and tenure matter to investors? Evidence from the bond market. *Journal of Accounting Research* 42 (4): 755–793.
- Marques, A. 2006. SEC interventions and the frequency and usefulness of non-GAAP financial measures. *Review of Accounting Studies* 11 (4): 549–574.
- Meyer, B. 1995. Natural and quasi-natural experiments in economics. *Journal of Business and Economic Statistics* 12: 151–162.
- Morsfield, S., and C. Tan. 2006. Do venture capitalists influence the decision to manage earnings in initial public offerings? *The Accounting Review* 81 (5): 1119–1150.
- Muller, K., and E. Riedl. 2002. External monitoring of property appraisal estimates and information asymmetry. *Journal of Accounting Research* 40 (3): 865–881.
- Nichols, C., J. Wahlen, and M. Wieland. 2009. Publicly traded versus privately held: Implications for conditional conservatism in bank accounting. *Review of Accounting Studies* 14 (1): 88–122.
- Omer, T., J. Bedard, and D. Falsetta. 2006. Auditor-provided tax services: The effects of a changing regulatory environment. *The Accounting Review* 81 (5): 1095–1117.
- Petersen, M. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22 (1): 435–480.
- Phillips, J. 2003. Corporate tax-planning effectiveness: The role of compensation-based incentives. *The Accounting Review* 78 (3): 847–874.
- Pincus, M., and S. Rajgopal. 2002. The interaction between accrual management and hedging: Evidence from oil and gas firms. *The Accounting Review* 77 (1): 127–160.
- Pittman J., and S. Fortin. 2004. Auditor choice and the cost of debt capital for newly public firms. *Journal of Accounting and Economics* 37 (1): 113–136.
- Puhani, P. 2000. The Heckman correction for sample selection and its critique. *Journal of Economic Surveys* 14 (1): 53–68.
- Rogers, J. 2008. Disclosure quality and management trading incentives. *Journal of Accounting Research* 46 (5): 1265–1296.
- Roulstone, D. 2003. The relation between insider-trading restrictions and executive compensation. *Journal of Accounting Research* 41 (3): 525–551.
- Ruddock, C., S. J. Taylor, and S. L. Taylor. 2007. Nonaudit services and earnings conservatism: Is auditor independence impaired? *Contemporary Accounting Research* 23 (3): 701–46.
- Schrand, C., and F. Wong. 2003. Earnings management using the valuation allowance for deferred tax assets under SFAS No. 109. *Contemporary Accounting Research* 20 (3): 579–611.
- Thursby, J. 1988. Evaluation of coefficients in misspecified regressions with application to import demand. *The Review of Economics and Statistics* 70 (4): 690–695.
- Tucker, J. 2007. Is openness penalized? Stock returns around earnings warnings. *The Accounting Review* 82 (4): 1055–1087.
- Tucker, J. 2010. Selection bias and econometric remedies in accounting and finance research. *Journal of Accounting Literature* 29: 31–57.
- Verrecchia, R., and J. Weber. 2006. Redacted disclosure. *Journal of Accounting Research* 44 (4): 791–814.
- Weber, J., and M. Willenborg. 2003. Do expert informational intermediaries add value? Evidence from auditors in microcap initial public offerings. *Journal of Accounting Research* 41 (4): 681–720.
- Wier, H. A. 2009. Fair value or conservatism: The case of the gold industry. *Contemporary Accounting Research* 26 (4): 1207–33.
- Zhang, H. 2009. Effect of derivative accounting rules on corporate risk-management behavior. *Journal of Accounting and Economics* 47 (3): 244–264.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.