SAMPLE SELECTION BIAS AND HECKMAN MODELS IN STRATEGIC MANAGEMENT RESEARCH

S. TREVIS CERTO, JOHN R. BUSENBARK, HYUN-SOO WOO, and MATTHEW SEMADENI*

Department of Management, W. P. Carey School of Business, Arizona State University, Tempe, Arizona, U.S.A.

Research summary: The use of Heckman models by strategy scholars to resolve sample selection bias has increased by more than 700 percent over the last decade, yet significant inconsistencies exist in how they have applied and interpreted these models. In view of these differences, we explore the drivers of sample selection bias and review how Heckman models alleviate it. We demonstrate three important findings for scholars seeking to use Heckman models: First, the independent variable of interest must be a significant predictor in the first stage of a model for sample selection bias to exist. Second, the significance of lambda alone does not indicate sample selection bias. Finally, Heckman models account for sample-induced endogeneity, but are not effective when other sources of endogeneity are present.

Managerial summary: When nonrandom samples are used to test statistical relationships, sample selection bias can lead researchers to flawed conclusions that can, in turn, negatively impact managerial decision-making. We examine the use of Heckman models, which were designed to resolve sample selection bias, in strategic management research and highlight conditions when sample selection bias is present as well as when it is not. We also distinguish sample selection bias, a form of omitted variable (OV) bias, from more traditional OV bias, emphasizing that it is possible for models to have sample selection bias, traditional OV bias, or both. Accurately identifying the type(s) of OV bias present is essential to effectively correcting it. We close with several recommendations to improve practice surrounding the use of Heckman models. Copyright © 2015 John Wiley & Sons, Ltd.

Selection bias is not well understood by practitioners. (Kennedy, 2006: 286)

INTRODUCTION

Empirical studies in strategy research rely on samples of observations that represent fractions of underlying populations. Biases may arise when researchers use samples instead of populations to test hypotheses. In particular, sample selection bias may occur when values of a study's dependent variable are missing as a result of another process (Greene, 2011; Sartori, 2003). The central objective of this article is to explore the drivers of sample selection bias and review how different analytical tools can correct it.

Scholars routinely describe the intuition of sample selection bias as requiring a two-stage approach (e.g., Wooldridge, 2010). Determining whether or not an observation in an overall population appears in its final representative sample is the first stage, and modeling the relation between the hypothesized dependent and independent variables in the final sample is the second stage. When an omitted variable (i.e., an unmeasured variable not included in a

model) creates a correlation between the error terms in these two stages, traditional techniques such as ordinary least squares (OLS) regression may report biased coefficient estimates. To resolve this potential bias, Heckman (1976) introduced the Heckman model, a two-step process for data analysis.[1]

To better understand how strategy scholars approach potential sample selection bias, we reviewed 63 articles appearing in the *Strategic Management Journal* (*SMJ*) between 2005 and 2014 that utilized Heckman models. In recent years, strategy scholars have employed Heckman models to study areas, including upper echelons and board membership (e.g., Quigley and Hambrick, 2012), diversification and M&A activity (e.g., Kim, Hoskisson, and Lee, 2015), executive compensation (e.g., Chen, 2015), capital market activity (e.g., Arikan and Capron, 2010), and competition and factor markets (e.g., Ndofor, Sirmon, and He, 2011).

Despite the significant growth in the use of Heckman models in strategy research (more than 700 percent over the last decade), we noted inconsistencies in how strategy scholars implemented and reported the results they derived. We also found that scholars often justified the use of Heckman models on the basis of concerns about endogeneity from a source other than sample selection. This is perhaps to be expected, however, as some econometrics textbooks list sample selection as a potential cause of endogeneity (e.g., Kennedy, 2006). This practice may cause some researchers to (mistakenly) equate the effects of sample selection bias with the effects of other sources of endogeneity. These discrepancies suggest that, as strategy scholars, we need a more rigorous understanding of (1) *how* sample selection bias varies across study conditions, (2) *when* sample selection bias affects statistical results, (3) *how* to apply Heckman models, and (4) *how* to account for effects resulting from sample selection bias versus other sources of endogeneity.

Accordingly, the first objective of this article is to explain *how* sample selection bias varies across study conditions. We explain that sample selection bias is the result of a special case of endogeneity, which we label sample-induced endogeneity. This

special case occurs when omitted variables create a correlation between the error term in the selection equation (i.e., the first stage of a study's statistical model) and the error term in the equation of interest (i.e., the second stage). We explain that the magnitude and direction of the bias depends on two factors: (1) whether the true relationship between the independent and dependent variable is positive or negative, and (2) whether the correlation between the error terms in the two stages is positive or negative.

To address our first objective, we create four figures (Figure 1a–d) to demonstrate that, in some cases, sample selection bias can lead researchers to find significant relationships that do not exist, or in other cases, it can lead researchers to fail to find significant relationships that do exist. Then, we report on the three studies in which we used simulations to address our other three objectives. Study 1 examines the conditions necessary to understand *when* sample selection processes bias results. We examine two factors related to the sample selection process: (1) the strength of the correlation between the error terms from first- and second-stage equations, and (2) the extent to which the independent variable of interest is related to the probability of an observation's entering the final sample. Our literature review revealed a great deal of confusion regarding the role of lambda. Our simulations illustrate that a significant lambda does not always denote sample selection bias. Specifically, our results indicate that traditional techniques (e.g., OLS) remain unbiased when the independent variable from the second stage is not also a significant predictor in the first stage.

Study 2 uses simulations to clarify *how* researchers should implement Heckman models. In this study, our findings illustrate that lambda can be insignificant even when sample selection bias exists, if model specifications are improper. This finding is important as our review uncovered numerous strategy research interpreting an insignificant lambda as evidence of no sample selection bias. Taken together, the simulations illustrate the complex role of lambda in explaining sample selection bias and help researchers understand how to use Heckman models.

Finally, Study 3 addresses *how* researchers should approach different sources of endogeneity when employing Heckman models. We distinguish between sample-induced and other forms of endogeneity. Our simulations reveal that Heckman

---

[1] By "the Heckman model," we are referring to the Heckman two-stage model. There is an alternative Heckman model known as the Heckman full information maximum likelihood (FIML) model, but the more ubiquitous is the two-stage model. Of the 63 reviewed articles published in the *SMJ* between 2005 and 2014, only one article used the Heckman FIML model.
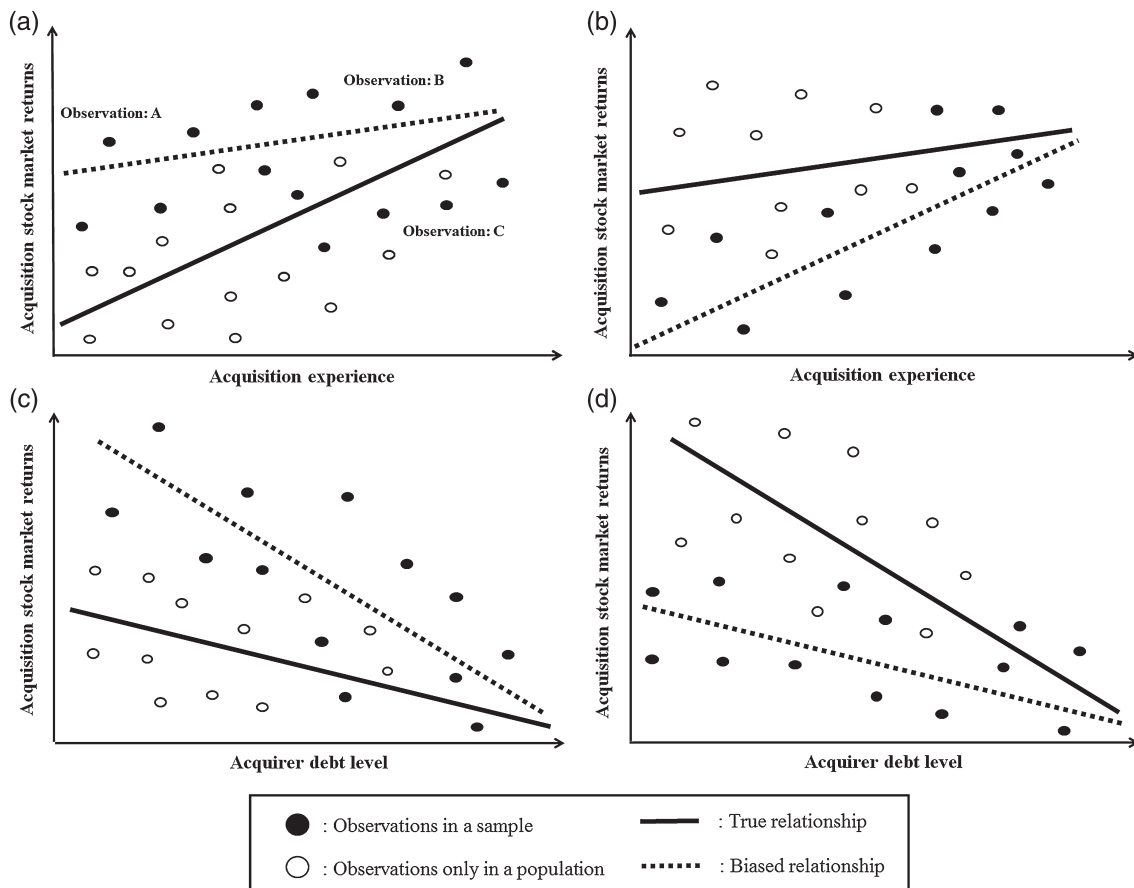
Figure 1.   (a) Positive beta and positive rho (Type II error). (b) Positive beta and negative rho (Type I error). (c) Negative beta and positive rho (Type I error). (d) Negative beta and negative rho (Type II error)

models help to resolve the endogeneity resulting from sample selection, but do not account for independent variables that are endogenous for other reasons. Our results help researchers to better understand how similar—yet different—analytical techniques (e.g., Heckman models, two-stage least squares) resolve different sources of endogeneity.

NONRANDOM SAMPLES AND SAMPLE SELECTION BIAS

Nonrandom samples

Econometricians and statisticians agree that the gold standard for any empirical study investigating causal relationships in the social sciences is random sampling and experimentation (e.g., Angrist and Pischke, 2008). In a random sample, the average characteristics of both observable and omitted variables should mirror those of the population

it represents (Vella, 1998). Researchers using a random sample can employ OLS regression to test the hypothesized relationship between the dependent and independent variable (or variables) of interest as captured by the following equation:

$$y_i = a + \beta x_i + e_i. \qquad (1)$$

Under the maintained assumption that errors (e) have a mean of zero, and are independent and identically distributed, OLS provides unbiased and efficient estimates of beta, which is the magnitude of the influence of independent variable(s) x on dependent variable y.

We focus on acquisition activity to illustrate the sample selection process.[2] Specifically, we

---

[2] Our example corresponds to Heckman's example of women in the workforce. To clarify, the relationship between acquisition experience (education level) and stock market reactions (wages)

review the role of acquisition experience in explaining stock market reactions to acquisition announcements (e.g., Haleblian & Finkelstein, 1999; Haleblian, Kim, & Rajagopalan, 2006). One approach to studying this relationship would involve using OLS and Equation 1 to estimate the effect (β) of acquisition experience (x) on stock market reactions (y) for a firm (i). With a random sample of firms, OLS would produce an unbiased and efficient estimate of this effect.

However, since stock market reactions are only available for firms that actually complete acquisitions, this research design necessarily leads to collection of a nonrandom sample: firms undertaking acquisitions. Firms undertaking acquisitions may differ from those that do not do so. For example, firms with extensive post-acquisition integration capabilities may undertake acquisitions at higher rates than firms with less extensive post-acquisition integration capabilities because these capabilities will help to improve post-acquisition performance (e.g., Zaheer, Castañer, and Souder, 2013; Zollo and Singh, 2004). In her review of sample selection methods, Sartori (2003) describes this process using the following utility function:

$$d_i = a + \beta w_i + u_i. \tag{2}$$

In Equation 2, d represents the utility of firm i's undertaking an acquisition; w represents a vector of variables that determine the likelihood of entering the sample, and u represents the errors that are independent and identically distributed with a mean of 0. As d increases, the probability of a firm's undertaking an acquisition increases. Instead of observing d, though, researchers only observe a dichotomous variable, D, which equals 1 when a firm undertakes an acquisition and 0 otherwise. This variable D occurs when the utility from undertaking an acquisition exceeds a specified threshold.

In econometrics textbooks, sample selection is typically tied to the idea of sample truncation (e.g., Baum, 2006; Kennedy, 2006; Wooldridge, 2010). Truncated samples exist when the values of the independent variable are unknown because the dependent variable is unobserved for part of the relevant population. As Wooldridge (2010) points

out, this often occurs when a researcher examines a subset of a population. For instance, a researcher may have data only on individuals with incomes of $75,000 or less per year.[3] Analyses based on such samples are not generalizable to the entire population.

Sample selection involves *incidental* truncation. Incidental truncation occurs when the dependent variable is "observed only if other variables take on particular values" (Wooldridge, 2010: 777). Extending our example of acquisitions, samples of acquisition announcements are based on the result of another process: the decision to engage in acquisition activity. For those firms not engaging in acquisition activity, there are no corresponding stock market reactions to acquisition announcements.

Sample selection bias

The sample selection process may introduce bias when OLS is used. Extending our example, a nonrandom sample of acquisitions is likely to include firms with relatively high acquisition experience. Sartori (2003) explains, however, that this likelihood by itself does not induce bias. If the sample only included these firms with more acquisition experience, the results would remain unbiased.[4] Instead, the potentially biasing factor is that this nonrandom sample also includes some firms with little or no acquisition experience that choose to undertake acquisitions despite their lack of experience. The firms with lower acquisition experience that are most likely to undertake an acquisition, and thus, enter the sample are those that have higher— and unobserved—post-acquisition integration capabilities, which offset their lack of experience. It is this omitted variable that may lead to bias.

A nonrandom sample potentially biases the results of OLS when an omitted variable influences both (1) the probability of entering the sample (e.g., acquiring another firm), and (2) the

---

depends on the decision to acquire another company (enter the workforce). A correlation among the error terms in the selection and structural equations may arise because of post-acquisition integration capabilities (intelligence).

[3] Truncated samples are similar, but not identical, to censored samples. Censored samples involve observations wherein the dependent variable is unobserved, but the independent variable is observed (Baum, 2006; Wooldridge, 2010). Extending this example, a censored sample may include the entire population, but all incomes *above* $75,000 are coded *as* $75,000. This is also known as "top-coding" (Wooldridge, 2010: 605).

[4] At the same time, though, examining the effects of only a subset of values of the independent variable may result in range restriction problems, which decrease efficiency and statistical power (Cohen *et al.*, 2003).

ultimate dependent variable of interest (e.g., stock market reaction). In such contexts, the omitted variable creates a correlation between the two error terms in Equations 1 and 2 (Kennedy, 2006; Sartori, 2003). In our example, the unmeasured variable "post-acquisition integration capabilities" is positively correlated with the error term, u, in the selection model (i.e., capabilities positively influence the probability of a firm's undertaking an acquisition) as well as the error term, e, in the regression equation (i.e., capabilities are positively associated with stock market reactions). This positive correlation suggests that the expected value of e in the final sample is positive and not equal to 0 as a result of sample selection bias, and so resulting estimates will be biased.

All else being equal, firms with low acquisition experience (x) have lower values of d, which is the dependent variable in the first stage. For firms with low values of x, the only way d can become large enough to warrant acquiring another firm is for the error term, u, to be unusually high. When u is correlated with the error term in the second stage (e), firms with low acquisition experience are also associated with larger values of e, which induces a negative correlation between x and e.

The graphs in Figure 1(a–d) illustrate the bias resulting from the sample selection process described above. Figure 1(a) plots acquisition experience on stock market reactions for a sample of firms undertaking acquisitions. The solid black line represents the hypothesized "true" relationship between acquisition experience and stock market reactions (e.g., Hayward, 2002). The filled circles represent firms in the final sample (i.e., observed observations), and the unfilled circles represent firms in the population that are not in the final sample (i.e., unobserved observations). The figure illustrates that the observed observations at lower levels of acquisition experience all appear above the solid line. This occurs because firms with less acquisition experience undertake acquisitions because they have greater post-acquisition integration capabilities (i.e., omitted variable). Because of these capabilities, they are more likely to undertake acquisitions and receive more positive stock market reactions (i.e., higher y values). In Figure 1(a), observation A, associated with a low value of x, represents an example of a large, positive residual with no offsetting negative residuals. In contrast, observations B and C illustrate that at higher levels of x observations exist both above and below the regression line.

The dotted line denotes the biased relationship between acquisition experience and stock market reactions found with this nonrandom sample. The higher residuals (e.g., observation A) on the left portion of the graph help to "lift" the left side of the line, which flattens the slope of the dotted line. If all firms—even those with less developed integration activities—undertook acquisitions, there would be negative residuals to negate these positive residuals, and the "true" relationship would appear. By ignoring this sample selection bias, however, OLS would report a coefficient corresponding to the slope of the dotted line as the relationship. As compared to the "true" relationship, though, the dotted line represents a *weaker* relationship between the two variables. In this case, then, using OLS would result in Type II errors (i.e., failures to detect real significant relationships).

The inverse is true when the relationship between x and y is positive and an omitted variable creates a negative correlation between e and u. Research suggests that the omitted variable "CEO hubris" might result in a negative correlation between e and u. CEOs with high levels of hubris may be more likely to conduct acquisitions, but acquisitions undertaken by these same CEOs may receive more negative stock market returns (e.g., Hayward and Hambrick, 1997).

Figure 1(b) illustrates the effects of a positive relationship between acquisition experience and stock market returns in the presence of a negative correlation between u and e. In this case, OLS reports a relationship that *exceeds* the true value. Stated differently, this combination results in Type I errors (i.e., false findings of significance).

Figure 1(c and d) illustrate that similar patterns emerge when we examine an independent variable that is negatively related to the dependent variable. Researchers (e.g., Finkelstein and Haleblian, 2002) have suggested that acquirer debt is negatively related to stock market reactions to acquisitions. Figure 1(c) shows the effects of a negative relationship between debt level and stock market returns in the presence of a positive correlation between u and e (resulting from the omitted variable, post-acquisition integration capabilities). In this situation, the true relationship is not as steep as the reported relationship, which results in Type I errors. Finally, Figure 1(d) illustrates the effects of a negative relationship between x and y (e.g., high-debt acquirers) and a negative correlation between u and e (e.g., resulting from the omitted

variable, CEO hubris). In such contexts, OLS will result in Type II errors.[5]

In summary, sample selection bias is not a condition that always inflates or deflates findings. Instead, sample selection bias depends on whether the relationship under investigation is positive or negative, and whether the correlation between the two error terms is positive or negative. Stated differently, sample selection bias is heterogeneous and varies with study conditions.

Sample-induced endogeneity versus other sources of endogeneity

Although sample selection bias may induce endogeneity, it is important to note that the correlation between x and e in a final sample may result from other sources, such as measurement error, autoregression, and simultaneous causality (e.g., Bascle, 2008; Semadeni, Withers, and Certo, 2014; Shaver, 1998). In strategy research models, endogeneity typically occurs because of an omitted variable that is correlated with both the independent variable and the error term (Bascle, 2008). Semadeni *et al.* (2014) demonstrate that even low levels of endogeneity can result in biased coefficient estimates.

This distinction has important implications for understanding the relationship between endogeneity and sample selection bias. Sample-induced endogeneity can result from one omitted variable, while traditional endogeneity can result from a different omitted variable or variables. At times, these two different types of endogeneity can create bias in the same direction. At other times, these different types of endogeneity can create bias in opposite directions. Accordingly, it is important for researchers to identify all sources of endogeneity before data analysis because different techniques resolve different types of endogeneity.

Addressing sample-induced endogeneity: Heckman's solution

Heckman introduced a two-stage process to correct sample-induced endogeneity. The first stage in this process uses a probit model (Equation 2) to estimate the probability of an observation's entering a sample, and the second stage uses OLS (Equation 1) to predict the ultimate dependent variable. To account for the potential biases that may result from nonrandomness, this process uses Equation 2 (in conjunction with Equation 1) to create a selection parameter, the inverse Mills ratio (IMR). This selection parameter is then included in Equation 1, where the coefficient is referred to as lambda, to account for potential sample selection bias. Lambda can be computed as sigma multiplied by rho, where sigma represents the standard deviation of the residuals in the second-stage equation, and rho is the correlation between error terms in the first- and second-stage equations.

Heckman models should include at least one variable in the first stage that does not appear in the second stage (Sartori, 2003). These variables, which are known as exclusion restrictions, influence the probability of an observation's appearing in the sample, but do not influence the ultimate dependent variable of interest in the second-stage OLS model. Without exclusion restrictions, Heckman models "can often do more harm than good" (Kennedy, 2006: 271).[6]

Wooldridge (2010) suggests that the effectiveness of this technique relies on meeting three assumptions. First, the independent variable of interest is available for the broader population, while the dependent variable is only available in the selected sample. Examining the role of the independent variable in both stages—and not just the second stage—of the analysis helps to check this assumption (Lennox, Francis, and Wang, 2011). Second, an omitted variable creates a correlation between the two error terms (i.e., e and u) in the selection and structural equations. Scholars can explore this assumption by theorizing about potential omitted variables that may lead to such a correlation. Empirically, researchers can either examine the correlation between e and u or report the significance of lambda, which is based on this correlation. Third, the independent

---

[5] Our simulations reveal another complexity in determining the direction of bias from sample-induced endogeneity; this involves the direction of x in the first-stage equation. In the examples here, we assume x remains positive in the first stage. In essence, this suggests that debt (financial leverage) is positively related to the probability of acquiring.

[6] The term *exclusion restriction* has been used in various ways. Some use the term to refer to the first-stage variables in Heckman models that do not appear in the second stage (e.g., Sartori, 2003). In contrast, some econometricians use the term more generally to describe the *property* of a variable that has a population coefficient of zero with another (e.g., dependent) variable (e.g., Wooldridge, 2010). Throughout this article, we use the term to refer to variables in the first stage of Heckman models that do not appear in the second stage.

variable is not correlated with either *e* or *u,* and is thus exogenous in both stages (Lennox *et al.*, 2011).

Together, these three conditions are necessary for employing Heckman models effectively. To better understand how strategy scholars utilize Heckman models, we reviewed published research. In the following discussion, we highlight how scholars employing Heckman models reported data related to the three noted assumptions.

HECKMAN MODELS IN STRATEGIC MANAGEMENT RESEARCH

We examined empirical research papers in *SMJ* between 2005 and 2014 to understand how strategy researchers justify and apply sample selection models. Heckman models were used in more than eight percent of the articles for this 10-year period. Moreover, the use of Heckman models in *SMJ* increased dramatically over time. In 2005, just under three percent of *SMJ* articles employed Heckman models. By 2014, approximately 25 percent of articles in this same journal employed them—an increase of more than 700 percent in 10 years.

The role of *x*

Regarding the first assumption involving the role of *x* in the first stage of modeling, we found that approximately 36 percent of the *SMJ* articles reported the findings from the first stage, and 41 percent reported the included exclusion restrictions, but only 27 percent reported the statistical significance of these exclusion restrictions. Nevertheless, almost two-thirds of the articles did not report the first stage, making it impossible to understand how these strategy researchers assessed whether or not observations for the independent variable were available when observations for the dependent variable were not.

Omitted variables in the two stages

We also examined the second assumption, which involves an omitted variable that influences both stages of the sample selection process by creating a correlation between the two error terms, *u* and *e*, in the two stages of Heckman models. In our review, only 35 percent of the articles stated that the

Heckman model was employed because of omitted variables that may have influenced either of the stages. Statistically, rho represents the correlation between *u* and *e*, and lambda is a function of rho. Lambda was reported in approximately 71 percent of the articles, and 78 percent of these identified the significance of lambda. Rho was only reported in approximately nine percent of the articles. In sum, a large number of articles in our review did not report the extent of the correlation between the two error terms, *u* and *e*.

The correlation between *x* and *e* and/or *u*

The third assumption involves an independent variable that is uncorrelated with either *e* or *u* (i.e., *x* is exogenous in both stages). This assumption seems to confuse scholars using Heckman models. More than 33 percent of the articles explicitly identified traditional endogeneity as the motivation for employing Heckman models, but we suspect that many more papers used Heckman models to correct for traditional endogeneity without specifically identifying it. This is concerning given that Heckman models were designed to resolve sample-induced endogeneity and not other forms of endogeneity.

SIMULATIONS

General setup

In light of the inconsistencies uncovered in our review of strategy research, we designed simulations to address three primary issues. First, how can researchers better understand when to employ Heckman models? Study 1 examines the extent to which the correlation between the error terms in the two stages of the sample selection process results in sample selection bias. Study 1 also investigates the role of *x* in the first stage of a Heckman model in explaining sample selection bias. Second, how should researchers evaluate the appropriateness of exclusion restrictions? Study 2 investigates how researchers might use inverse Mills ratios and pseudo-R^2 measures to evaluate exclusion restriction strength. Finally, to what extent do other sources of endogeneity influence the results of Heckman models, which were developed to account for sample-induced endogeneity? Study 3 examines the relative effectiveness of Heckman models and

two-stage least squares models in remedying these different sources of bias.

Simulations involve two primary steps: (1) generating datasets with realistic and known parameters, and (2) analyzing the datasets with different models to better understand how the results match with the true parameters. To create simulations that represent realistic scenarios, we relied on existing research on effect sizes (Cohen, 1992; Cohen *et al.*, 2003), previous simulations (Semadeni *et al.*, 2014), and our review of strategy research using Heckman models. The simulations in our three studies share general data generation characteristics and outcome variables, which we detail in the following discussion.

Data generation

Simulating a population effect

To establish true relationships among variables, we generated two datasets: the total population of 1,000 observations (N) and a nonrandom sample of 500 observations (n) selected from this population. We chose these values to approximate the median sample size we observed in our analysis of the *SMJ* articles. A further description of the data generation process can be found in Appendix S1.

Our objective was to model a small effect in the population relationship between the independent and dependent variable. We used Equation 1 to generate the relationship between x and y for observation i in the population N. Effect size is a function of several factors, including the variance of the independent variable (x), the variance of the random error term (e), and β. We generated N observations of x (mean of 1 and variance of 1) and e (mean of 0 and variance of 2) (Semadeni *et al.*, 2014). We set β equal to 0.15 for all simulations. Using these conditions, x accounts for approximately two percent of the variance in y in the OLS model using the total population, which approximates Cohen's (1992) criteria for small effects in the context of OLS regression (see also Cohen *et al.*, 2003).[7]

Simulating sample selection

After establishing the relationship between x and y in the total population, we introduced sample selection using Equation 3:

$$d_i = \alpha + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \gamma_3 z_{3i} + \gamma_4 x_i + u_i. \quad (3)$$

In this equation, the selection parameter (d) is a function of three exclusion restrictions (z_{1i}, z_{2i}, and z_{3i}), the independent variable of interest (x), and a random error term (u) for each observation (i). The random error term (u) was normally distributed with a mean of 0 and a variance of 2, which is consistent with assumptions of the probit model and also equals the value of the variance in the second-stage error term, e, in Equation 1.

Much has been written about model fit in probit models (e.g., Hosmer, Hosmer, Le Cessie, and Lemeshow, 1997; Long and Freese, 2006), particularly about pseudo-R^2 as a metric. We included different exclusion restriction coefficients in order to approximate small, medium, and large effects (Cohen, 1992), using McFadden pseudo-R^2 values as a guide. Specifically, we set γ_1, γ_2, and γ_3 to 0.5, 1, and 1.5, respectively. These coefficients result in individual pseudo-R^2 values of approximately 0.02, 0.10, and 0.20, respectively. When we set γ_4 to 0.5, combined with the exclusion restrictions, it resulted in an average pseudo-R^2 value of approximately 0.35. In the absence of exclusion restrictions, x represents a pseudo-R^2 of approximately 0.02, which corresponds with Cohen's (1992) definition of a small effect size.[8] The dependent variable in the first stage, d, represents a continuous latent variable. When values of d exceeded a specified threshold, D was set to 1 (i.e., the observation was included in the selected sample), and 0 otherwise (i.e., the observation was not included in the selected sample).[9]

Data analysis

Estimators

We compared and contrasted three different estimation models. First, to understand the "true"

---

[7] As a robustness check, we also ran simulations with control variables. The results for these simulations were substantively similar. We should highlight that the results changed when the control variables induced multicollinearity, but such problems with multicollinearity are not specific to Heckman models and exist even with less sophisticated techniques such as OLS. For simplicity, we omit control variables from our reported simulations.

[8] It is important to note that we varied these parameters in a comprehensive sensitivity analysis, and our results were substantively similar.

[9] With "real" datasets, researchers observe 0s and 1s. Here, we generate a latent variable that is converted to a 1 or 0 if it exceeds a specific threshold. Our data use a threshold of 50 percent, wherein the observation appears as 1 if the latent variable is greater than the mean of the latent variable values. We altered this using several different levels and did not find substantive differences.

relationship between x and y, we employed OLS regression *on the population* using Stata's "*-regress-*" command. We label this technique and its results *Population OLS*. Second, we performed OLS regression on the incidentally truncated sample. This technique, which we label *Incidentally Truncated OLS* (henceforth, *I.T. OLS*), allowed us to better understand the bias that results from using OLS regression on nonrandom samples.

Third, we used the Heckman two-stage procedure on the incidentally truncated sample, labeling technique and results *Heckman Model*. We used Stata's "*-heckman, twostep-*" command. For the baseline conditions used in Studies 1 and 3, we included only the strongest exclusion restriction in the Heckman model. In other words, we generated the data using Equation 3 to include multiple possible exclusion restrictions, but in the analysis we assumed that we could access only one of the exclusion restrictions, reflecting a realistic research scenario.

Outcome variables

We employed four general outcome variables to better understand how sample selection issues influence the bias and efficiency of different estimators. $Beta_{Avg}$ represents the average beta reported across the 1,000 iterations of a simulation.[10] The extent to which beta departs from the true value in the total population indicates bias. SE_{Avg} represents the average standard error reported in the simulations. All else held equal, lower SE_{Avg} values denote increased efficiency. $PerSig_x$ represents the percentage of simulations, wherein the reported coefficient for x is statistically significant at the 95 percent level. Finally, $PerSig_\lambda$ represents the percentage of the simulations in which the lambda is statistically significant at the 95 percent level.

STUDY 1: WHEN SAMPLE SELECTION BIAS OCCURS

In our simulations examining the conditions necessary for sample selection bias, correlation between the error terms in the first (selection) and second (estimation) stages of a sample selection process is the first parameter. The second is the extent to which

the independent variable (x) of interest in the second stage influences the likelihood of selection in the first stage. We varied both of these parameters to better understand sample selection bias and the models used to circumvent this bias.

Simulation conditions

First, we varied the correlation between the error terms in the first stage (u) and the second stage (e) to correspond to no (corr[e,u] = 0), low (corr[e,u] = 0.10), and moderate (corr[e,u] = 0.30) correlations (Semadeni *et al.*, 2014). Second, we varied the extent to which x predicted the probability of entering the final sample by altering its coefficient in the first stage. Specifically, we created three conditions by varying γ_4, which represents the relationship between x and d in Equation 3. In line with the coefficients we chose for the exclusion restrictions in Equation 3, we created conditions where x had no effect ($\gamma_4 = 0$), a small effect ($\gamma_4 = 0.50$), or a medium effect ($\gamma_4 = 1$) in determining the probability of entering the sample. With the exception of these two conditions (i.e., the corr[e,u] and the coefficient [γ_4] for x in the first stage), all of the other parameters remained constant and as described in our "General setup" section.

Results

Table 1 illustrates the outcome measures for Study 1. Panel A, which involves models with a medium effect size of x in the first stage, shows the effects of an increasing correlation between the error terms, u and e. When u and e are uncorrelated, $Beta_{Avg}$ remains consistent across all three models. As the correlation between u and e increases, though, the results of the *I.T. OLS* model become more biased. $Beta_{Avg}$ for *I.T. OLS* is 0.13 when the correlation between u and e is 0.10, and $Beta_{Avg}$ decreases to 0.08 when the correlation increases to 0.30. Furthermore, in these cases $PerSig_x$ is 50.15 and 23.52 percent, respectively.

Comparing Panels A, B, and C illustrates how the effect size of x in the first stage influences sample selection bias. Panel B, where x has a small effect size, illustrates a pattern similar to Panel A, yet more moderate. Panel C, which displays the results when x has no effect in the first stage, illustrates a different effect. Specifically, even when the correlation between the error terms increases to

---

[10] For all of the variables, we also calculated median values. The differences between the mean and median values are minimal.

Table 1.   When sample selection bias occurs

Rho condition	Parameter	I Population OLS	II I.T. OLS	III Heckman Model
Panel A—X medium effect in the first stage				
Corr(e,u) = 0	*BetaAvg (True value = 0.15)*	0.15	0.15	0.15
	SEAvg	0.04	0.07	0.08
	PerSig$_x$ (True value = 95%) (%)	92.29	63.56	51.55
	PerSig$_\lambda$ (%)			3.90
Corr(e,u) = 0.1	*BetaAvg (True value = 0.15)*	0.15	0.13	0.15
	SEAvg	0.04	0.07	0.08
	PerSig$_x$ (True value = 95%)(%)	93.09	50.15	54.15
	PerSig$_\lambda$ (%)			10.01
Corr(e,u) = 0.3	*BetaAvg (True value = 0.15)*	0.15	0.08	0.15
	SEAvg	0.04	0.07	0.07
	PerSig$_x$ (True value = 95%) (%)	90.89	23.52	48.95
	PerSig$_\lambda$ (%)			45.25
Panel B—X low effect in the first stage				
Corr(e,u) = 0	*BetaAvg (True value = 0.15)*	0.15	0.15	0.15
	SEAvg	0.04	0.06	0.07
	PerSig$_x$ (True value = 95%) (%)	90.39	66.67	62.36
	PerSig$_\lambda$ (%)			4.40
Corr(e,u) = 0.1	*BetaAvg (True value = 0.15)*	0.15	0.14	0.15
	SEAvg	0.04	0.06	0.07
	PerSig$_x$ (True value = 95%)(%)	92.19	60.46	63.46
	PerSig$_\lambda$ (%)			10.41
Corr(e,u) = 0.3	*BetaAvg (True value = 0.15)*	0.15	0.12	0.15
	SEAvg	0.04	0.06	0.07
	PerSig$_x$ (True value = 95%) (%)	92.09	44.34	64.36
	PerSig$_\lambda$ (%)			50.95
Panel C—X no effect in the first stage				
Corr(e,u) = 0	*BetaAvg (True value = 0.15)*	0.15	0.15	0.15
	SEAvg	0.04	0.06	0.06
	PerSig$_x$ (True value = 95%) (%)	90.99	63.76	63.96
	PerSig$_\lambda$ (%)			4.90
Corr(e,u) = 0.1	*BetaAvg (True value = 0.15)*	0.15	0.15	0.15
	SEAvg	0.04	0.06	0.06
	PerSig$_x$ (True value = 95%) (%)	91.99	67.67	68.37
	PerSig$_\lambda$ (%)			8.61
Corr(e,u) = 0.3	*BetaAvg (True value = 0.15)*	0.15	0.15	0.15
	SEAvg	0.04	0.06	0.06
	PerSig$_x$ (True value = 95%) (%)	92.29	68.67	67.87
	PerSig$_\lambda$ (%)			51.35

moderate levels (i.e., corr[e,u] = 0.30), no sample selection bias arises when x is not a significant predictor in the first stage.[11] Both *Beta$_{Avg}$* and *PerSig$_x$* remain generally unchanged across all of the conditions in Panel C of Table 1. Taken together, the results in Panels A, B, and C demonstrate the two necessary conditions for sample selection bias: x must be a significant predictor in the first stage,

and a correlation between the error terms in the two equations must exist.

Panel A in Table 1 also illustrates how the significance of lambda changes as the correlation between u and e increases. When the correlation between u and e is 0, *PerSig$_\lambda$* is 3.90 percent. When the correlation is 0.30, however, *PerSig$_\lambda$* increases to over 45.25 percent. These figures are consistent regardless of the effect size of x in the first stage (i.e., across Panels A, B, and C). Although many scholars use the significance of lambda as an indicator of sample selection bias, our results suggest that a

---

[11] Although statistical power is lower in the *I.T. OLS* than the *Population OLS* because the sample size is half, we see that the *Heckman Model* and the *I.T. OLS Model* are nearly identical.

significant lambda alone does not indicate sample selection bias. Moreover, the simulations highlight that it is difficult to assess sample selection bias on the basis of lambda alone.

Discussion

In brief, the findings from Study 1 demonstrate that there are two necessary conditions for sample selection bias: x must be a significant predictor of the first stage *and* the error terms e and u must be correlated. Using the statistical significance of lambda as a proxy for sample selection bias is only appropriate if x is a significant predictor in the first stage. Some scholars in econometrics have questioned the use of the significance of lambda as an indicator of sample selection bias (e.g., Guo and Fraser, 2009), and our results support this perspective. To our knowledge, very few scholars have explored the role of x in the first stage. Our findings suggest that x must be a significant predictor of an observation's being included in the final sample for sample selection bias to occur. If x is not significant in the first stage, though, sample selection will not emerge—regardless of a correlation between error terms e and u.

In addition, our results indicate that a nonzero correlation between e and u is a necessary condition for sample selection bias. When there is no correlation between these terms, *I.T. OLS* is superior to *Heckman Model*. Even at low levels of correlation between the error terms, however, sample selection bias emerges. This correlation also has implications for the statistical significance of lambda. Regardless of whether or not sample selection bias is present, lambda becomes significant more frequently as the correlation between the error terms increases.

STUDY 2: HOW TO USE THE HECKMAN TWO-STAGE MODEL

In Study 1, we demonstrated that two conditions are necessary for sample selection bias to be present. In this study, we turn our attention to how researchers should employ Heckman models to correct sample selection bias. Heckman models require exclusion restrictions in the first stage that do not appear in the second stage. In this study, we examine how the strength of exclusion restrictions alters the parameters reported by Heckman models.

Our review of the literature suggests that strategy scholars typically conflate exclusion restrictions in Heckman models with instruments in two-stage least squares. While conceptually similar, exclusion restrictions differ from instruments in two ways. First, exclusion restrictions are exogenous variables that predict whether or not an observation appears in a sample (Angrist, 2001). Instrumental variables, on the other hand, are exogenous variables intended to represent endogenous independent variables (Hamilton and Nickerson, 2003; Semadeni *et al.*, 2014; Stock, Wright, and Yogo, 2002). Second, instruments differ from exclusion restrictions in that instruments substitute for the endogenous variable in a second-stage estimation. In contrast, Heckman models incorporate exclusion restrictions to compute an adjustment factor (i.e., IMR) that is included in the second-stage estimation (Angrist, 2001; Bushway, Johnson, and Slocum, 2007). In both cases, the variables chosen (whether an instrument or exclusion restriction) should not correlate with the error term associated with the dependent variable in the second stage (e).

While exclusion restrictions pertaining to Heckman models have generated much discussion (e.g., Bushway *et al.*, 2007; Leung and Yu, 1996; Sartori, 2003), there is little consensus regarding the assessment of the appropriateness of exclusion restrictions. Bushway *et al.* (2007: 164) note that because it is difficult to define the properties of good exclusion restrictions, they must be determined on "substantive rather than technical grounds." Nevertheless, some scholars have proposed evaluating the strength of exclusion restrictions by examining the correlation between the inverse Mills ratio and the independent variable, x (Bushway *et al.*, 2007; Leung and Yu, 1996; Moffitt, 1999). When the exclusion restrictions are poor in a model (or do not exist at all), IMR will correlate too highly with x, thus introducing multicollinearity problems in the second stage of the model. In this study, we examine this correlation as well as the pseudo-R^2 associated with the first stage to see if these two measures yield any insights regarding exclusion restriction strength.

Simulation conditions

Because we are concerned with the strength of exclusion restrictions in this study, we focus on the first stage of a Heckman model, the selection equation. To briefly review our simulation design,

we used three different coefficients for three different exclusion restrictions in Equation 3 to generate the data. In an ideal scenario, a researcher would include all three of these variables, which would perfectly specify the Heckman model. However, it is unrealistic that a researcher would find all of the potential exclusion restrictions.

In this study, we created four conditions that could occur when a researcher employs a Heckman model. Specifically, the researcher could have access to no exclusion restrictions, an exclusion restriction with a small effect ($\gamma_2 = 0\ 0.5$), one with a medium effect ($\gamma_2 = 1$), or one with a large effect ($\gamma_3 = 1.5$). Examining these four different conditions allowed us to better understand how exclusion restriction strength influences the outcomes associated with Heckman models.[12]

We deviated slightly from the general setup used in the previous study in two ways. First, because we were examining the effectiveness of exclusion restrictions when sample selection bias is present, we constrained the correlation between the error terms to a medium level of 0.30. Second, because we were only examining the efficacy of Heckman models in this study, we report two additional outcomes associated with them. Specifically, following research on exclusion restriction strength (Bushway *et al.*, 2007; Leung and Yu, 1996), we report the correlation between *x* and IMR, labeled *corr[x,IMR]*. We also report the pseudo-R^2 associated with the first stage.

Results

Table 2 displays the results of Study 2. Columns I–IV illustrate how the Heckman model estimates vary as exclusion restriction strength increases. As shown in Columns II–IV, *Beta*$_{avg}$ remains fairly consistent when the exclusion restrictions are nonzero. In contrast, the standard errors (*SE*$_{avg}$) decrease as exclusion restriction strength increases. This combination had direct implications for *PerSig*$_x$, which increases as exclusion restriction strength increases.

Table 2 also illustrates how stronger exclusion restrictions influence the correlation between *x* and IMR as well as the pseudo-R^2. Specifically, as the strength of the exclusion restriction increases (i.e.,

its coefficient increases), *corr[x,IMR]* decreases.[13] When there is no exclusion restriction in the model, *corr[x,IMR]* approaches 1. This correlation decreases to 0.69 when the exclusion restriction is weakest and decreases further to 0.31 with the strongest exclusion restriction. In a similar pattern, the pseudo-R^2 associated with the first stage increases as exclusion restriction strength increases (ranging from 0.02 to 0.24). Taken together, both of these measures illustrate the potential effectiveness of exclusion restrictions.

Table 2 also illustrates the limitations of using the significance of lambda as an indicator of sample selection bias. In each of these conditions, sample selection bias exists. The ability to detect a significant lambda, though, depends on exclusion restriction strength. For example, *PerSig*$_\lambda$ is approximately seven percent with the weakest exclusion restriction, but increases to approximately 51 percent with the strongest exclusion restriction.

Figure 2(a and b) illustrate the implications of our results. Together, these figures illustrate how the correlation between *x* and IMR influences the extent to which *x* and lambda are statistically significant. We include various sample sizes because of the influence of sample size on statistical power (Cohen, 1992; Cohen *et al.*, 2003). Together these figures illustrate that stronger exclusion restrictions influence the statistical significance of both *x* and lambda.

Discussion

The results contained in Study 2 help provide two contributions regarding exclusion restrictions and sample size. First, our results suggest that the correlation between *x* and IMR provides a useful indicator of exclusion restriction strength. Second, we demonstrate how exclusion restriction strength—in conjunction with sample size—influences the statistical significance of both the independent variable and lambda in the second stage.

We propose that researchers can use the correlation between *x* and IMR as an indicator of exclusion

---

[12] We include all four exclusion restrictions in the data generation process to ensure stable pseudo-R^2 values across conditions.

[13] The correlations stated are in absolute values. In actuality, the correlation between *x* and lambda are negative because sample selection bias results in a higher correction factor at lower levels of *x*. If *x* is negatively related in the first stage, then this correlation would be positive.

Table 2.   Effects of the varying exclusion restriction strength

Outcome variable	I *No ER*	II $\gamma = 0.5$	III $\gamma = 1$	IV $\gamma = 1.5$
BetaAvg (True value = 0.15)	0.14	0.15	0.15	0.15
SEAvg	5.49	0.09	0.07	0.07
PerSig$_x$ (True value = 95%) (%)	0.00	38.94	55.16	64.36
PerSig$_\lambda$ (%)	0.00	6.61	19.82	50.95
Corr[x,IMR]	1.00	0.69	0.44	0.31
Pseudo R^2	0.02	0.04	0.11	0.24

The correlation between *x* and the *inverse Mills ratio* is stated in absolute terms. Depending on the direction of the effect of x in the first stage, this value can be positive or negative, but is interpreted the same in either circumstance.

restriction strength. We hesitate to provide exact values of that correlation at which exclusion restrictions are "strong enough," because sample size and other factors (such as the correlation between error terms) would change any benchmark values. Nevertheless, we suggest using this value in conjunction with other elements of a Heckman model to assess the strength of the exclusion restrictions and efficacy of the model in general. For example, as rho (or the correlation between *e* and *u*) increases, so too will the statistical significance of lambda. When interpreting lambda, scholars should also examine rho to look for any substantive correlation between the error terms.[14] Further, scholars can also review the pseudo-R^2 values associated with the first stage to better understand exclusion restriction strength, where larger values represent stronger exclusion restrictions.

Our study also demonstrates that an insignificant lambda may not indicate an absence of sample selection bias. If a sample is small and/or exclusion restrictions are weak, Heckman models are unlikely to produce significant lambdas—even in the presence of sample selection bias. Combined with the results in Study 1, Study 2's results caution against using either significant or insignificant lambdas as an indicator of whether or not sample selection bias exists. Nevertheless, the Heckman model does appear to produce unbiased coefficient estimates, even when exclusion restrictions are weak. Consequently, we expect scholars will find accurate estimates even with weak exclusion restrictions, but such cases will likely result in inaccurate standard errors.

---

[14] One can calculate the statistical significance of rho using its value and the sample size (Cohen, 1992).

STUDY 3: HECKMAN MODELS AND THE SOURCE OF ENDOGENEITY

Studies 1 and 2 examined when sample selection bias occurs and how a Heckman model can resolve it. Nevertheless, our review of the literature suggested that some strategy scholars employ Heckman models to resolve endogeneity emanating from sources other than sample selection. Kehoe and Tzabbar (2014: 719), for example, state: "We corrected for potential endogeneity by using a two-stage Heckman selection procedure." Similarly, Yang, Lin, and Lin (2010: 249) state that they used the inverse Mills ratio from a Heckman model to "avoid potential bias due to endogeneity." We highlight these two examples, but many other examples exist where scholars employ a Heckman model without specifying sample-induced endogeneity.

In this study, we investigate how Heckman models account for different sources of endogeneity. In Study 1, we modeled a correlation between *u* and *e*. This correlation created sample-induced endogeneity when *x* was a determinant of the model's first stage. The primary difference between the first study and the current one is that here, we also created a direct correlation between *x* and *e*, which reflects the type of endogeneity typically discussed in strategy research (e.g., Semadeni *et al.*, 2014).

Extending the example of the relationship between acquisition experience and announcement effects may help to explain the distinction between different sources of endogeneity. With sample-induced endogeneity, acquisition experience should be exogenous (i.e., not correlated with the error term in the second stage in a model). Suppose, for example, that an omitted variable (e.g., post-acquisition integration capability) is
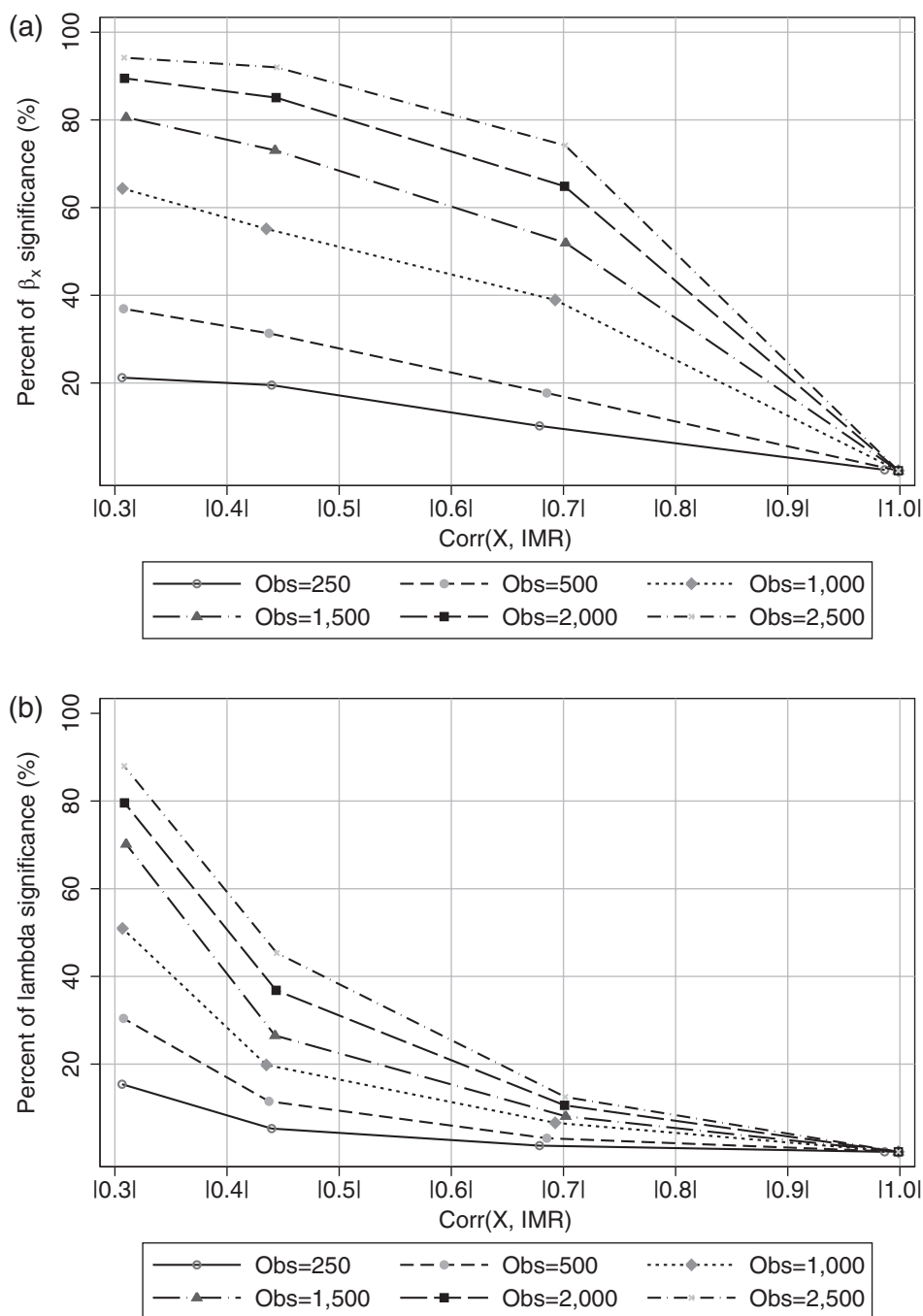
Figure 2.   (a) Exclusion restriction strength (percent of *βx* significant). (b) Exclusion restriction strength (percent of *lambda* significant). Notes: (1) "Obs" corresponds to the population observations (*N*). (2) rho = 0.3. (3) We used the absolute values of Corr(X, IMR) for a better graphical presentation

correlated with acquisition experience as well as announcement returns. In this circumstance, acquisition experience would be endogenous because it is correlated with the error term in the second stage (*e*).

Heckman models were not designed to correct for omitted variables that are also correlated with an independent variable (Wooldridge, 2010). As such, how do Heckman models perform when traditional endogeneity and sample-induced endogeneity are

both present? To address that question, in this third simulation, we created both traditional endogeneity and sample-induced endogeneity in a single model.

Simulation conditions

The simulation conditions for Study 3 include the parameters discussed in "General setup" with two exceptions. First, we also model an endogenous x by varying the correlation between x and e to equal 0 or 0.10 in the population (e.g., Semadeni *et al.*, 2014). Second, we use two-stage least squares (*2SLS*) models, a conventional technique used to resolve traditional endogeneity, to analyze the data, and we include an instrument, z, which was correlated with x at 0.30 (e.g., Semadeni *et al.*, 2014).[15]

Results

The results of this simulation are displayed in Table 3. Panel A replicates previous simulations in that we only model sample-induced endogeneity. Panel B reports results when the correlation between x and e increases to 0.10. Column III of Panel A illustrates that when the correlation between e and u is 0, the Heckman model reports $Beta_{Avg}$ values of 0.15. In contrast, Panel B (where the correlation between x and e increases to 0.10) illustrates that the Heckman model reports $Beta_{Avg}$ values of approximately 0.29, even when there is no sample-induced endogeneity.

As illustrated in Column IV of Table 3, the *2SLS* model greatly reduces the bias created by endogeneity, even when sample selection bias is present. Panel B illustrates that when x is correlated with e, *2SLS* is more accurate than the other estimators. In addition, *2SLS* (Column IV) reports estimates that are also less likely to be statistically significant than estimates reported by other models. 2SLS models, then, are less biased than Heckman models when traditional endogeneity and sample-induced endogeneity are both present. Still, though, here the *Heckman Model* is less biased than *2SLS* when only sample-induced endogeneity exists, and not traditional endogeneity (as demonstrated in Panel A).

---

[15] For robustness, we also used two instruments. The results are effectively equivalent whether one or two instruments are used. The *F* associated with the first-stage varies, but is always substantially greater than 10, which is the minimum suggested benchmark for strong instruments (Stock *et al.*, 2002).

Discussion

The results of this study have important implications for strategy researchers. Heckman models account for only sample-induced endogeneity and do not resolve other types of endogeneity. Our simulations suggest that if x and e are even mildly correlated, Heckman models will report biased results. Although our results highlight the benefits of Heckman models for exogenous independent variables, our results also highlight that 2SLS models are more appropriate than Heckman models when both types of endogeneity are present. Unfortunately, in much of the current literature, scholars appear to conflate the two types of endogeneity.

RECOMMENDATIONS

We provided several contributions through our review of the sample selection process, our review of the implementation of Heckman models in strategy research, and our simulations. We clarified how sample selection arises and highlighted the heterogeneity in sample selection bias. We also noted several inconsistencies among strategy scholars using Heckman models. To investigate the implications of these inconsistencies, we created a series of simulations involving sample selection and Heckman models.

Although we struggled to identify a single paper that implemented a Heckman model perfectly, the complexities of sample selection bias and Heckman models perhaps underscore the difficulty of the task. Nevertheless, we believe that our work contributes a series of recommendations for scholars investigating the potential role of sample selection in their research. In the following sections, we highlight a number of steps that researchers can take when confronting potential sample selection.

Consider the potential for sample selection

The first step in implementing a Heckman model involves considering the potential for sample selection bias. As we noted previously, our literature review suggests that scholars often conflate sample selection bias with endogeneity from other sources. A simple rule helps to distinguish between these two alternatives: If the dependent variable is

Table 3.   Traditional versus sample-induced endogeneity

Rho condition	Parameter	I Population OLS	II Truncated OLS	III Heckman Model	IV 2SLS
Panel A — No endogeneity: Corr(x,e) = 0					
Corr(e,u) = 0	*BetaAvg (True value = 0.15)*	0.15	0.15	0.15	0.17
	SEAvg	0.04	0.06	0.07	0.13
	PerSig$_x$ (True value = 95%) (%)	90.40	64.60	60.60	26.00
	PerSig$_\lambda$ (%)			5.60	
Corr(e,u) = 0.1	*BetaAvg (True value = 0.15)*	0.15	0.14	0.15	0.17
	SEAvg	0.04	0.06	0.07	0.13
	PerSigx (True value = 95%) (%)	92.80	60.80	64.60	24.20
	PerSig$_\lambda$ (%)			9.00	
Corr(e,u) = 0.3	*BetaAvg (True value = 0.15)*	0.15	0.11	0.15	0.14
	SEAvg	0.04	0.06	0.07	0.13
	PerSigx (True value = 95%) (%)	91.20	43.20	63.20	18.40
	PerSig$_\lambda$ (%)			50.20	
Panel B — Low endogeneity: Corr(x,e) = 0.1					
Corr(e,u) = 0	*BetaAvg (True value = 0.15)*	0.29	0.29	0.29	0.18
	SEAvg	0.04	0.06	0.07	0.13
	PerSigx (True value = 95%) (%)	100.00	99.60	99.40	29.40
	PerSig$_\lambda$ (%)			4.40	
Corr(e,u) = 0.1	*BetaAvg (True value = 0.15)*	0.29	0.28	0.29	0.16
	SEAvg	0.04	0.06	0.07	0.13
	PerSigx (True value = 95%) (%)	100.00	99.00	98.80	24.60
	PerSig$_\lambda$ (%)			9.80	
Corr(e,u) = 0.3	*BetaAvg (True value = 0.15)*	0.29	0.26	0.29	0.13
	SEAvg	0.04	0.06	0.07	0.13
	PerSigx (True value = 95%) (%)	100.00	98.20	99.40	18.60
	PerSig$_\lambda$ (%)			54.60	

observed for only a subsample of a larger population (e.g., wages for working women, stock market reactions to acquisition announcements, etc.), sample selection is a potential problem. In such cases, Heckman models (i.e., Stata's "-*heckman*-") may help to resolve potential sample selection bias. In contrast, if the dependent variable is available for all observations (e.g., ROI of acquirers versus nonacquirers), then sample selection is likely not an issue. Instead, in such cases the independent variable is likely endogenous, which would require either two-stage least squares (i.e., Stata's "-*ivreg*-") or treatment effects models (i.e., Stata's "-*etregress*-").

Consider the role of omitted variables

If the potential for sample selection is confirmed, we suggest using Figure 1(a–d) to better understand how it may bias statistical results. These figures demonstrate that both the direction and magnitude of sample selection bias depend on the relationship between the independent and dependent variables as

well as the nature of the correlation between *u* and *e*, the error terms in the two equations. In some cases, sample selection bias will result in Type I errors, while in other instances, the bias will result in Type II errors.

Given this heterogeneity, we recommend that researchers carefully consider how sample selection bias may arise in their studies. Which omitted variable(s) may be producing a correlation between the error terms? Is the potential correlation positive or negative? In addition, we urge reviewers to abandon the practice of simply noting "I think this paper suffers from sample selection problems" and moving on to the next point. We admit that, at times, we have provided such comments to authors, but the complexity of potential sample selection bias illustrates that this practice should cease. Instead, we suggest that reviewers recommend to authors specific omitted variables and corresponding anticipated effects. Our simulations illustrate the counterintuitive point that, in some cases, OLS provides conservative estimates of relationships in the presence of sample selection.

Evaluate role of *x* in first stage

One of our central contributions in this study involves the role of *x* in the first stage of a model. To our knowledge, no previous study has explicitly highlighted the important role *x* plays in sample selection. We recommend that scholars investigating potential sample-induced endogeneity first observe the role of *x* in the first stage of their estimation. If *x* is not significant in the first stage of a Heckman model, our results suggest that sample selection bias will not exist. In such instances, we suggest scholars use a different estimator (e.g., OLS) to test their hypotheses, because sample-induced endogeneity will not create bias.

Evaluate the significance of lambda

If *x* is significant in a first stage, we recommend employing a Heckman model and examining the significance of lambda in the model's second stage. If *x* is significant in the first stage and lambda is significant in the second, scholars may report the results from Heckman models. It is important to note that observing a significant lambda in the second stage does not denote sample selection bias. For example, Quigley and Hambrick (2012: 843) find that lambda "was significant … suggesting some bias was present." While this is one of the few articles we reviewed that recognizes the relevance of a significant lambda, sample selection bias requires both a significant *x* in the first stage and a significant lambda. Because information provided by lambda's statistical significance incorporates rho (i.e., the correlation between *e* and *u*), we recommend that researchers focus more on lambda than rho when evaluating potential sample selection.

If lambda is insignificant in the second stage of a model, however, we caution against dismissing potential sample selection bias. For example, Rubera and Tellis (2014: 2050) note that lambda "turned to be insignificant, suggesting that there is no selection bias in our sample." Our simulations indicate that weak exclusion restrictions and/or small samples may result in insignificant lambdas, even when sample selection bias is present. Our simulations also suggest that scholars should examine the correlation between *x* and the inverse Mills ratio to better understand exclusion restriction strength.

Figure 2(a and b) provide some general guidance regarding how the significance of both *x* and lambda fluctuate as the correlation between *x* and IMR increases. These figures also demonstrate that such relationships vary depending on the sample size. If the correlation between *x* and IMR is too high, or the pseudo-R^2 is too low, we recommend that scholars search for stronger exclusion restrictions. Although researchers can use our figures to approximate exclusion restriction strength, we should note that the figures represent our specific simulation conditions and may not generalize to all situations. While we contribute to the literature by providing an overview of these relationships, we are hopeful that future research can further explore specific guidelines regarding exclusion restriction strength.

Consider other sources of endogeneity

Our final contribution involves examining the effectiveness of Heckman models when endogeneity arises from sources other than sample selection bias. Our simulations suggest that other sources of endogeneity bias statistical results more than sample-induced endogeneity. We also find that Heckman models perform poorly when other, more traditional, sources of endogeneity exist. Heckman models were designed to account for sample-induced endogeneity only, not for other sources of endogeneity. Moreover, when both sample-induced and alternately sourced endogeneity exist, we find that two-stage least squares models provide less biased estimates than Heckman models. Researchers should only employ Heckman models to resolve potential sample selection (i.e., there is a correlation between *e* and *u*). In contrast, we recommend 2SLS or treatment effects models to resolve endogeneity from other sources that create a correlation between *x* and *e*. When research contexts include both sample selection bias and endogeneity from other sources, we recommend using and comparing both Heckman and 2SLS models. If the two models report substantively different results, we urge researchers to favor 2SLS over Heckman. Simply stated, our simulation results suggest that researchers should be more concerned with endogeneity from other sources than with sample selection bias.

CONCLUSION

Understanding both *when* and *how* to use Heckman models is complex and involves several moving

pieces. Taken together, our review of the broader literature on sample selection and our simulations help clarify the heterogeneity of sample selection bias, the different types of endogeneity, and some procedures useful to circumvent these problems. We present practices for identifying the conditions that produce sample selection bias as well as techniques necessary for resolving it.

ACKNOWLEDGEMENTS

REFERENCES

Angrist JD. 2001. Estimation of limited dependent variable models with dummy endogenous regressors. *Journal of Business and Economic Statistics* **19**(1): 2–28.

Angrist JD, Pischke J-S. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press: Princeton, NJ.

Arikan AM, Capron L. 2010. Do newly public acquirers benefit or suffer from their pre-IPO affiliations with underwriters and VCs? *Strategic Management Journal* **31**(12): 1257–1289.

Bascle G. 2008. Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization* **6**(3): 285–327.

Baum CF. 2006. *An Introduction to Modern Econometrics Using Stata*. Stata Press: College Station, TX.

Bushway S, Johnson BD, Slocum LA. 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.

Chen G. 2015. Initial compensation of new CEOs hired in turnaround situations. *Strategic Management Journal* **36**(12): 1895–1917.

Cohen J. 1992. A power primer. *Psychological Bulletin* **112**(1): 155–159.

Cohen J, Cohen P, West SG, Aiken LS. 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd edn). Lawrence Erlbaum Associates: Mahwah, NJ.

Finkelstein S, Haleblian J. 2002. Understanding acquisition performance: the role of transfer effects. *Organization Science* **13**(1): 36–47.

Greene WH. 2011. *Econometric Analysis*. Prentice Hall: Englewood Cliffs, NJ.

Guo S, Fraser MW. 2009. *Propensity Score Analysis: Statistical Methods and Applications*. Sage Publications: Thousand Oaks, CA.

Haleblian J, Finkelstein S. 1999. The influence of organizational acquisition experience on acquisition performance: a behavioral learning perspective. *Administrative Science Quarterly* **44**(1): 29–56.

Haleblian JJ, Kim J-yJ, Rajagopalan N. 2006. The influence of acquisition experience and performance on acquisition behavior: evidence from the US commercial banking industry. *Academy of Management Journal* **49**(2): 357–370.

Hamilton BH, Nickerson JA. 2003. Correcting for endogeneity in strategic management research. *Strategic Organization* **1**(1): 51–78.

Hayward ML. 2002. When do firms learn from their acquisition experience? Evidence from 1990 to 1995. *Strategic Management Journal* **23**(1): 21–39.

Hayward ML, Hambrick DC. 1997. Explaining the premiums paid for large acquisitions: evidence of CEO hubris. *Administrative Science Quarterly* **42**(1): 103–127.

Heckman JJ. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* **5**(4): 475–492.

Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. 1997. A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine* **16**(9): 965–980.

Kehoe RR, Tzabbar D. 2014. Lighting the way or stealing the shine? An examination of the duality in star scientists' effects on firm innovative performance. *Strategic Management Journal* **36**(5): 709–727.

Kennedy P. 2006. *A Guide to Econometrics*. MIT Press: Cambridge, MA.

Kim H, Hoskisson RE, Lee SH. 2015. Why strategic factor markets matter: "new" multinationals' geographic diversification and firm profitability. *Strategic Management Journal* **36**(4): 518–536.

Lennox CS, Francis JR, Wang Z. 2011. Selection models in accounting research. *Accounting Review* **87**(2): 589–616.

Leung SF, Yu S. 1996. On the choice between sample selection and two-part models. *Journal of Econometrics* **72**(1): 197–229.

Long JS, Freese J. 2006. *Regression Models for Categorical Dependent Variables using Stata*. Stata Press: College Station, TX.

Moffitt R. 1999. New developments in econometric methods for labor market analysis. In *Handbook of Labor Economics*, Ashenfelter OC, Card D (eds). North-Holland: Amsterdam, Netherlands; 1367–1397.

Ndofor HA, Sirmon DG, He X. 2011. Firm resources, competitive actions and performance: investigating a mediated model with evidence from the in-vitro diagnostics industry. *Strategic Management Journal* **32**(6): 640–657.

Quigley TJ, Hambrick DC. 2012. When the former CEO stays on as board chair: effects on successor discretion, strategic change, and performance. *Strategic Management Journal* **33**(7): 834–859.

Rubera G, Tellis GJ. 2014. Spinoffs versus buyouts: profitability of alternate routes for commercializing

innovations. *Strategic Management Journal* **35**(13): 2043–2052.

Sartori AE. 2003. An estimator for some binary-outcome selection models without exclusion restrictions. *Political Analysis* **11**(2): 111–138.

Semadeni M, Withers MC, Certo ST. 2014. The perils of endogeneity and instrumental variables in strategy research: understanding through simulations. *Strategic Management Journal* **35**(7): 1070–1079.

Shaver JM. 1998. Accounting for endogeneity when assessing strategy performance: does entry mode choice affect FDI survival? *Management Science* **44**(4): 571–585.

Stock JH, Wright JH, Yogo M. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* **20**(4): 518–529.

Vella F. 1998. Estimating models with sample selection bias: a survey. *Journal of Human Resources* **33**(1): 127–169.

Wooldridge JM. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA.

Yang H, Lin Z, Lin Y. 2010. A multilevel framework of firm boundaries: firm characteristics, dyadic difference, and network attributes. *Strategic Management Journal* **31**(3): 237–261.

Zaheer A, Castañer X, Souder D. 2013. Synergy sources, target autonomy, and integration in acquisitions. *Journal of Management* **39**(3): 604–632.

Zollo M, Singh H. 2004. Deliberate learning in corporate acquisitions: post-acquisition strategies and integration capability in US bank mergers. *Strategic Management Journal* **25**(13): 1233–1256.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix S1. Stata code