Endogeneity and Entrepreneurship Research

Remedies Using Instrument Variables & Two Stage Least Squares

Brian S. Anderson^{*†}

Abstract

In this paper, I discuss common endogeneity problems found in entrepreneurship research. I show how entrepreneurship researchers, particularly those working with observational data, typically face several specific endogeneity threats in a given study. I then show through a series of simulated datasets instrument variable approaches that effectively address endogeneity in a variety of research designs. I discuss best-practice recommendations for integrating instrument variables into entrepreneurship research at the study design phase, and outline potential sources for instruments. To facilitate replication and to provide a guide for scholars, I posted all data and code, in R and in Stata, online at: <https://osf.io/d453n/>.

1 Introduction

Central to our role as social scientists is accuracy—having confidence that the results we report from our empirical analysis are, to the best possible extent, free from bias (Bettis, 2012). As the replication crisis in science grows, fields from medicine (Ioannidis, 2016), to psychology (Lindsay, 2017), to management (Bergh et al., 2017) are looking inward and acknowledging the challenges inherent to identifying, and reporting, accurate empirical findings. The desire to have each published paper make a novel—and ideally important—contribution to a theoretical conversation creates a perverse incentive to find results that are first and foremost ‘interesting’ (Butler et al., 2017; Goldfarb and King, 2016). The failure to replicate several high-profile—interesting—phenomenon exposed the tension between the desire to publish novel findings and the long, often painstaking process necessary to demonstrate the robustness of those findings (Lindsay, 2017; Nuijten et al., 2016).

Entrepreneurship research certainly qualifies as interesting (Shepherd, 2015). The meteoric rise in the popularity of entrepreneurship scholarship testifies to the theoretical—and practical—importance of entrepreneurship phenomenon (Shepherd, 2015). Reflecting the topic we study, entrepreneurship researchers are opportunity-driven, in the sense that we eagerly discover, and create, new opportunities to push the entrepreneurship research frontier. This opportunity-driven spirit infuses vigor into our scholarship, which is critically important to the development of the field. Yet as with our sister disciplines, vigor and being interesting

^{*}I would like to thank Jeff Covin, Anand Titus, Matt Semadeni, Nik Franke, Jeff McMullen, and seminar participants at the University Nebraska, the University of Southern California, and the University of Missouri, Kansas City, for their helpful comments on earlier versions on this paper. Corresponding author: andersonbri@umkc.edu

[†]University of Missouri, Kansas City & Ghent University, Belgium

are of limited use without accuracy; that is, vigor without rigor is vapid (Bergh et al., 2017). We have the opportunity to continue the remarkable evolution of entrepreneurship scholarship—and the value we provide to our stakeholders—by continuing to infuse rigorous methods with the fundamentally interesting questions we ask.

The central premise of this paper is that internal validity—the ability to draw causal inference from a study—is a necessary, although not sufficient, condition for usefulness for studies testing theory. I purposefully draw a distinction between exploratory research and theory testing research. Studies offering a causal hypothesis—entrepreneurial intentions positively influence entrepreneurial behavior, for example—carry a higher burden for causal inference than studies offering exploratory analyses of an entrepreneurship phenomenon. In order for a theory-testing paper to be both interesting and useful, it must first be accurate, or as Antonakis (2017, p. 6) noted, “research that is not rigorous simply cannot be relevant. . .”. Establishing internal validity requires eliminating alternative explanations for the relationship of interest; in other words, dealing with the omitted variable problem. Omitted variables stem from selection effects, measurement error, simultaneity, an unmodeled common cause, or just bad luck (Antonakis et al., 2010). A study must address these threats, which manifest as endogeneity in a statistical model, in order to recover consistent—accurate—model estimates (Rubin, 2005).

In this paper, I discuss significant endogeneity threats to entrepreneurship research. I focus on a particular research question in entrepreneurship—the extent to which entrepreneurship education and training programs influence the probability of starting a new business—and delineate where endogeneity arises in a study on this representative topic. I then demonstrate across a series of commonly employed research models—a direct effect model, a mediation model, a latent variable model, and a moderation model—how one method to deal with multiple sources of endogeneity, instrument variables/Two Stage Least Squares (2SLS), allows a researcher to recover consistent parameter estimates.

Before proceeding, it is important to specify the target audience for this paper. Entrepreneurship research, broadly speaking, is arguably ‘late to the game’ in terms of dealing with endogeneity and related threats to causal inference. While entrepreneurship scholars are becoming more aware of the endogeneity problem, there remains significant room for improvement, particularly for researchers using regression, ANOVA/MANOVA, maximum likelihood, and structural equation modeling estimators, in both observational and experimental designs.

My paper makes three contributions to entrepreneurship research. The first contribution is identifying common endogeneity problems faced by entrepreneurship scholars and how to address the problem. My hope is that with this understanding, entrepreneurship researchers embrace identifying endogeneity problems during the research design phase of a study (Angrist and Pischke, 2008). The long-term goal is for entrepreneurship scholars to make solving the identification problem—recovering consistent estimates of the effect under study—as central to designing empirical work as decisions about operationalizations, sampling, and other critical factors. I also hope that the discussion aids reviewers and editors of entrepreneurship manuscripts to evaluate the endogeneity threats, and remedies, in a particular paper, and to provide guidance to authors to strengthen the causal inference of submitted papers.

The paper’s second contribution is to illustrate how an instrument variable/2SLS approach integrates into commonly found research models and effectively addresses the endogeneity problem. 2SLS is not the only approach to dealing with endogeneity—regression discontinuity, propensity score analysis, and difference-in-differences models being among several viable alternatives (Antonakis et al., 2014). However, 2SLS is a commonly found tool in the management literature (Semadeni et al., 2014), and is easily employed in a variety of applied research settings, making it a particularly robust method for recovering consistent parameter estimates. I demonstrate how the instrument variable/2SLS estimator effectively addresses endogeneity in main effect models, latent variable models, and mediation models, and provide a discussion of best practices for using 2SLS in moderation models.

The final contribution is an online repository for applied researchers. I posted all data and code, in R and in Stata, on the *Open Science Foundation* at <https://osf.io/d453n/>. I welcome and encourage researchers to download the data and code to reproduce and extend the models and techniques reported in the paper. I hope that these resources will be valuable for scholars in planning future studies and as templates for conducting analyses.

2 Understanding Endogeneity

2.1 Endogeneity Defined

Figure 1: Endogeneity and Consistency of Parameter Estimates

Endogeneity is a violation of the orthogonality assumption that the covariance between a predictor, x , and the disturbance term in a regression equation, ϵ , equals zero (Semadeni et al., 2014). Consider the graphic in Panel 1 of Figure 1. Here I depict a simple main effect model and no assumed covariance between x and ϵ ($\text{COV}[x, \epsilon] = 0$) and a simple linear equation:

$$y = \alpha + \beta x + \epsilon \quad (1)$$

In this model, the estimate of β will be consistent; β will converge asymptotically to the population value as the sample approaches infinity. In Panel 2, I show another model with a second predictor, z , that correlates with x and estimate the model with the following equation:

$$y = \alpha + \beta_1 x + \beta_2 z + \epsilon \quad (2)$$

Here I properly included z in the model, so estimates of x and z will both be consistent. If instead we were to estimate the model in Panel 3 with Eq. 1, $y = \alpha + \beta x + \epsilon$, our estimate of βx would be inconsistent, because the ‘true’ model remains $y = \alpha + \beta_1 x + \beta_2 z + \epsilon$. The dotted line in the figure indicates that z is a ‘true’ predictor of y , but we did not include it in the estimated model. Because z is a theoretically—and empirically—meaningful predictor of y , and assuming that x and z covary in practice, excluding z from the model will cause x to correlate with the disturbance term, ϵ .

To illustrate, consider a model where entrepreneurial self-efficacy, x , predicts start-up intentions, y . Assuming there are no other causes of entrepreneurial intentions that also correlate with entrepreneurial self-efficacy, we will get a consistent estimate of the effect of self-efficacy and intentions. Prior entrepreneurial experience, z , however, is also a meaningful predictor of entrepreneurial intentions (Krueger, 1993), and is known to covary with entrepreneurial self-efficacy (Zhao et al., 2005). Excluding prior entrepreneurial experience will bias the estimate of βx . When the estimator, be it regression, maximum likelihood, or other related approaches, solves for βx it will adjust β to satisfy the orthogonality assumption— βx is now biased, or inconsistent. Now matter how large the sample becomes, βx will always be wrong. The stronger the covariance between x and ϵ the greater the bias in βx . Importantly, the bias could be positive, amplifying the effect, or negative, attenuating the effect, depending on the covariance between x , and z , and y (Angrist et al., 1996).

2.2 Randomization and Endogeneity

Dealing with endogeneity is relatively simple, if the researcher can identify z . The challenge is that rarely, if ever, do we observe z , or are fully aware of all possible z ’s. In a sense, z is a vector of random variables representing all of the alternate explanations of the observed relationship between x and y that the researcher does not, or can not measure (Angrist and Pischke, 2008). The only way to ensure that this vector of unobservables does not systematically bias the observed relationship is to randomly assign study participants to a treatment condition for x (Antonakis et al., 2010). With random assignment, the researcher assumes that there is an approximately equal occurrence of each random z in the sample—for every participant with prior entrepreneurial experience there is a participant without—such that the unobservable z ’s in the model have a mean of zero (Angrist and Pischke, 2008).

Further, because the researcher controls, or manipulates x , the researcher assumes that there are no systematic differences (i.e., self-selection) in the value x takes for each participant (Antonakis et al., 2014). In the self-efficacy example, under random assignment, the researcher

would analogously be giving each participant a “high self-efficacy” pill or a “low self-efficacy” pill at random. Participants with high and low levels of entrepreneurial self-efficacy respectively before random assignment have an equal probability of receiving the self-efficacy manipulation (Rubin, 1974). When used effectively, this equivalent probability allows the researcher to safely ignore the unobservable z 's and the regression (or ANOVA, etc.) will yield a consistent, causal estimate of the effect of x on y . The observed variance in y may only be from the change in x made by the researcher.

2.3 Endogeneity as an Omitted Variable Problem

Unlike in a randomized controlled experiment, in an observational design, z is impossible to ignore (Rubin, 2005). An observational design is any study in which the researcher does not manipulate x , or the researcher does not randomly assign participants to a treatment condition. Surveys, secondary data, and experiments in which the respondent self-selects into the treatment condition are all observational designs (Imai et al., 2008). In these studies, there is no mechanism to ensure that a random cross-section of the population receives the treatment. Further, in an observational design, the observed value of x is under the control of the participant, not the researcher. As such, the researcher can not, and must not, assume that the vector of unobservable z 's in the model does not relate, in some way, with the variables that were observed by the researcher (Rubin, 2005).

In its simplest form, an omitted variable, z , is another variable accounting for the true relationship between x and y . Returning to $y = \alpha + \beta_1 x + \beta_2 z + \epsilon$ (Eq. 2), if we omit z , and if z correlates with x and with y , then the estimator yields an inconsistent value for βx . The reason is because we can think of variance in z as a function of variance in x :

$$z = \gamma x + \omega \quad (3)$$

Substituting Eq. 3 into our ‘true’ model for Eq. 2 yields:

$$y = \alpha + \beta_1 x + \beta_2 (\gamma x + \omega) + \epsilon \quad (4)$$

Multiplying out and rearranging yields:

$$y = \alpha + x(\beta_1 + \beta_2 \gamma) + (\beta_2 \omega + \epsilon) \quad (5)$$

No matter how large the sample, with z omitted, the effect of x and y will always be a partial function of x 's relationship with z . Importantly, if $\beta_2 = 0$ in Eq. 5, implying that the unobserved z does not cause y , Eq. 5 reduces to Eq. 1, $y = \alpha + \beta x + \epsilon$, yielding a consistent estimate of x . For omitted variable bias to be present, z must covary with x **and** with y .

Because endogeneity only manifests when the omitted variable covaries with x and y , it is helpful to think about endogeneity as impacting the relationship between x and y , and how

the researcher tests and interprets that relationship. This makes dealing with endogeneity not simply a statistical matter, but also a theoretical matter. To understand the possible omitted variables impacting the relationship between x and y , the researcher must have a deep understanding of the theoretical connection between x and y and the context in which the relationship manifests. In practice, this means that there is little utility to the statement that ‘ x is endogenous.’ The more accurate statement is that ‘ x may be endogenous to y because of the z unobserved/unobservable factors that are pertinent to this model.’

We can think of all endogeneity problems as a form of omitted variable bias in the context of a given empirical model (Antonakis et al., 2010). Selection effects, measurement error, common methods variance, simultaneity, and alternate causal explanations manifest as omitted z ’s—represented with Eq. 5 or a derivative. But the source of these omitted z ’s vary as a function of the conceptual meaning of x and y , how the data was collected, whether the researcher manipulated x , the accuracy of the measures, and so forth, as we will explain in more detail in Section 3. The endogeneity problem in one study will not necessarily be the endogeneity problem in another study, even with the same x and y .

2.4 Endogeneity and Control Variables

Revisiting Eq. 2, $y = \alpha + \beta_1 x + \beta_2 z + \epsilon$, we can easily infer that if we were to have identified z , we could control for its influence in our model. Theoretically, if we could identify all possible z ’s, and had a large enough sample (number observations > number of estimated parameters), we could include them as controls/covariates in a simple multiple regression model and recover a consistent estimate for βx . The problem, as we noted earlier, is that it is simply not possible to observe all possible omitted variables, and a correlation between ϵ and x may occur by random bad luck such that no control variable exists (Rubin, 2005). Conversely, and as we will discuss in Section 3.1, theoretically if the researcher identified the ‘perfect’ instrument for x , there would be no need to include any control variables, because the instrument would isolate the ‘true’ effect of x on y in an instrument variable/2SLS estimator. In practice, because we cannot observe all possible z ’s, nor is there a ‘perfect’ instrument for x , it remains a best practice to collect and include all theoretically meaningful controls/covariates. Being cognizant of model parsimony (Atinc et al., 2012), the researcher should limit covariates to those identifiable z ’s that are most likely, based on prior research and the researcher’s theoretical knowledge, to be confounding the relationship between x and y .

3 Endogeneity Challenges in Entrepreneurship Research

Because we can think of all endogeneity problems as a form of omitted variable bias (Antonakis et al., 2010), it is tempting to simply categorize observational and even experimental research as ‘having an endogeneity problem.’ This catch-all is tempting for reviewers and editors to

cite as justification for rejecting a paper, but is generally not helpful for researchers to develop an appropriate remedy. As such, it is helpful to understand the different ways in which an omitted z threatens the validity of entrepreneurship research, and the specific challenges facing entrepreneurship researchers to model causal effects, which is a topic we turn to next.

Consider a study estimating whether Entrepreneurship Education and Training (EET) programs increase the probability of starting a new business (Martin et al., 2013). The intuitive experimental research design, given a large cross-section of the population, would be to assign one group to an EET program, and another group to a control condition (no EET program). The difference in the probability of starting a business between the groups would be the assumed causal impact, or *average treatment effect*, of the EET program (Rubin, 2005). We can assert a causal relationship of EET on the probability of starting a new venture if our research design met the randomization requirement in Section 2.2. We observe, in this case, the *counterfactual*—the number of businesses started for a group that did not participate in the EET program.

The counterfactual, critical to causal reasoning, asks the question of what would we have observed if a given person that actually received the experimental treatment had hypothetically not received the treatment (Rubin, 2005, 1974). Random, controlled experiments allow us to observe the counterfactual not for an individual—a person cannot participate in an EET program and not participate at the same time—but for sample populations. Assuming differences between individuals occur at random across these samples, and that there are no factors influencing whether an individual received the treatment or not, we can safely ignore potential omitted variables and estimate a causal effect (Angrist and Pischke, 2008). Unfortunately, even in the case of random controlled experiments, there are threats to our ability to establish a causal relationship (Antonakis et al., 2010). For example, the researcher cannot compel an individual to participate in the experiment, a participant may drop out of the EET program before the experiment finishes, or a participant in the control condition may join a different EET program during the experiment.¹

There are several observational (non-experimental) ways to test whether participating in an EET program increases the probability of starting a new business (Martin et al., 2013). For illustration, we will use a telephone survey design, similar to the PSED/PSED II and related databases frequently used by entrepreneurship researchers. In our design example, we might call a large cross-section of individuals, employ a defined screening criteria (e.g., respondents must be working age, and so forth), and then ask whether at some point the respondent participated in an EET program. We then ask whether the respondent started a business or not after participating in the EET program. Using this design as our starting point, we next walk through four specific endogeneity threats, also referred to as identification problems, that may bias the estimated effect of EET programs on the probability of starting a new business. The four threats presented here are not exhaustive, but are commonly found in entrepreneurship research. Further, these threats would apply to *any* observational research design; they are not specific to telephone or Internet-based surveys.

¹Cook and Campbell (1979), among many others, is an excellent resource for understanding internal validity in experimental and quasi-experimental designs.

3.1 Selection Effects

Selection effects manifest in two forms: sample selection, and self-selection (Certo et al., 2016). Sample selection involves some type of sample truncation in which the researcher does not observe a random sample of the full population (Certo et al., 2016). Self-selection is non-random assignment into the treatment or measurement condition (Clougherty et al., 2016). Both types of selection effects frequently occur in observational research, and can be difficult to tease apart. In our EET survey design, we asked the respondent if he or she attended an EET program before starting a new venture, and so established *temporal sequencing* (x must come before y to establish causality). However, what may also be likely is that the respondent was already interested in or even planning to start a new business before attending the EET program, and simply attended the program to be better prepared. In this case, the selection effect would positively bias the estimate of attending an EET program on the probability of starting a new business. Non-response bias is also a form of selection effect in survey designs, where the researcher only observes those responding to the survey request.

In entrepreneurship research broadly, selection effects are a particularly salient problem (Clougherty et al., 2016). Individuals self-select into being entrepreneurs, firms purposely choose whether to adopt an entrepreneurial strategic posture, and managers dictate the risk-taking culture of the business. The survivor bias problem is another form of selection effect in entrepreneurship research (Wiklund and Shepherd, 2011), in which researchers generally cannot observe failed entrepreneurs or failed businesses. Importantly, while selection effects often inflate effect sizes (Shaver, 1998), it is also possible, depending on the research question, that a selection effect diminishes an estimated parameter (Certo et al., 2016).

3.2 Simultaneity

A common misconception associates endogeneity with reverse causality—the relationship between x and y could also be represented as y to x . While it is possible that there is causal variance in x that is due to changes in y (reverse causality), the more likely problem is simultaneity (Angrist and Krueger, 2001). Simultaneity is best thought of as the chicken-and-egg problem—it is inordinately challenging to derive the true causal sequence between x and y (Pearl, 2014). In the case of our EET programs illustration, we may observe a higher probability of starting a new business in areas where there are several EET programs. For example, we would expect to find multiple EET programs—incubators, accelerators, and mentorship networks—in geographic clusters known for startup activity (Roundy et al., 2017). As the region experiences more startup activity, however, it is reasonable to expect more EET programs, and hence more EET program participants. It is conceptually, and empirically, difficult to determine whether the EET program *caused* the observed startup activity, or whether the startup activity resulted in greater *a priori* participation in EET programs.

Another example of simultaneity in the entrepreneurship literature is the relationship between Entrepreneurial Orientation (EO) and firm growth. Growing the firm by behaving entrepreneurially involves entering new product/market spaces, which requires tangible and intangible resources (Anderson and Eshima, 2013). Growth, in turn, provides new resources

that enable future entrepreneurial behavior (Eshima and Anderson, 2017). Sorting out which came first—growth or entrepreneurial behavior—is difficult at best. This is why, for example, a technique such as ‘swapping’ the independent and dependent variables and evaluating whether the original dependent variable is a statistically significant predictor of the original independent variable yields little insight into whether endogeneity is present in a model. Regardless of whether modeled as a predictor or criterion, simultaneously determined variables will always be endogenous (Wooldridge, 2010).

3.3 Confounding Effects

This is the unknown-unknowns problem. The researcher specifies what he or she believes to be a true underlying relationship, but ‘misses’ identifying the actual variable that connects x and y together (Rubin, 2005). No matter how many times the researcher collects data on x and y , the relationship between them will always be wrong, because he or she did not collect, or more likely was not aware of, the confounding variable that causes both x and y . We can also think of a confounding effect as an alternative explanation—a reason for the observed variation in y that is also causing the observed variation in x , but that the researcher either does not, or cannot, directly observe (Antonakis et al., 2010).

Consider the relationship between EET programs, the probability of starting a new venture, and the entrepreneurial self-efficacy of the respondent (Zhao et al., 2005). It may be, for example, that individuals with high self-efficacy are more willing to both start a business, and to attend an EET program to enhance his or her entrepreneurial skill-set before starting the business. It could also be the case that an individual with low entrepreneurial self-efficacy chose to attend an EET program to evaluate whether he or she might be successful as an entrepreneur (Zhao et al., 2005). The EET program in turn boosted self-efficacy, which resulted in a greater probability of starting a business than would have happened had the respondent’s self-efficacy not have been low before starting the EET program. We could also envision a scenario in which a respondent had high perceived self-efficacy when entering an EET program, but after learning more about the challenges involved in business creation, chose *not* to start a business he or she was already planning because of a resulting decrease in self-efficacy.

The challenge with confounding effects is that there exists almost infinite alternative explanations for a given observed relationship. This is always the case for observational research, but also extends to experimental designs irrespective of random assignment into the treatment condition. Demand effects, participant drop-out, and participant apathy are potential confounds, but so too are weak or minimally effective experimental manipulations (Cook and Campbell, 1979). Experimental designs in which the researcher measures, but does not manipulate, a mediator variable also results in a confounding effect problem (Imai et al., 2010). Returning to our earlier experimental design example for EET programs, we might posit that entrepreneurial self-efficacy mediates the relationship between EET program participation and the probability of starting a new business. We might measure self-efficacy before and after the respondent took part in the EET program, and then model the change in self-efficacy as our mediator. The challenge is that while the causal mechanism could be

self-efficacy, because we did not manipulate the self-efficacy condition, it is equally possible that another variable, say self-esteem, is the ‘true’ mediator. With a confounding variable, even with theoretically relevant control variables, the researcher cannot rule out an alternate explanation for the relationship.

3.4 Measurement Error

A substantial portion of entrepreneurship research employs latent constructs. Because we cannot directly observe a latent construct—for example, entrepreneurial self-efficacy as the mediator in our EET program design—a researcher uses psychometric or other indicators meant to capture the underlying phenomenon (MacKenzie et al., 2011). Proxy variables also fall into this category (Boyd et al., 2013); the researcher collects an observed measure, research and development expenses for example, that he or she uses to represent a latent phenomenon (e.g., innovativeness, absorptive capacity, and so forth). Because we do not directly observe the latent construct (ξ), we represent its variance as a function of the assumed true score (λ_t), and random and systematic error ($\delta r, s$) (Podsakoff et al., 2003):

$$\xi = \lambda_t + \delta r, s \quad (6)$$

Measurement error will always introduce bias into effect size estimates. Substituting Eq. 6 for x in Eq. 1 yields:

$$y = \alpha + \beta(\lambda_t + \delta r, s) + \epsilon \quad (7)$$

Multiplying out and rearranging yields:

$$y = \alpha + \beta\lambda_t + \beta(\delta r, s + \epsilon) \quad (8)$$

As is clear from Eq. 8, the estimate of β will always be inconsistent so long as $\delta r, s$ does not equal zero.

Researchers often assess measurement error by evaluating the reliability of the construct’s indicators (e.g., Cronbach’s alpha). The assumption being that we can infer validity—that the latent construct captures the phenomenon of interest—through reliability. Importantly, higher reliability only *implies* lower measurement error and hence validity, because validity itself is not observable in a latent construct (MacKenzie et al., 2011). Further, even highly reliable constructs still retain some measurement error, leading to inconsistent parameter estimates. If the researcher collapses the indicators into a single summed or mean-scaled score, he or she assumes that measurement error equals zero ($\delta r, s = 0$ in Eq. 6). The reality for these models, however, is that $\delta r, s$ can never equal zero (Bollen, 1989), and so measurement error in the latent variable manifests as endogeneity (Antonakis et al., 2010).

Measurement error is also possible even with an observed secondary indicator that the researcher is *not* using to represent a latent phenomenon. For example, the researcher may be

modeling the relationship between spending on research and development and sales growth rate. A firm may, however, manipulate research and development expenses for earnings management or tax purposes rather than representing the ‘true’ amount spent on R&D activities (Boyd et al., 2013). Further, firms may employ different interpretations of Generally Accepted Accounting Principles, introducing measurement error into the data (Boyd et al., 2013). The net result is that researchers face a high bar in showing that measurement error is not materially present in his or her data, whether employing latent constructs or not.

3.5 Conclusion

It is often the case that an entrepreneurship study has several endogeneity threats. With observational data in particular, it is likely that all four issues discussed previously are present (Angrist and Pischke, 2008). Even in experimental designs, particularly those modeling mediation, endogeneity may exist from measurement error, selection effects, and confounding variables (Imai et al., 2010; Pearl, 2014). Escaping endogeneity is, then, not simply a research design issue nor simply a statistical correction issue. One advantage of the instrument variable (IV)/2SLS approach as we will discuss, however, is that when done well the method largely addresses endogeneity from each of the preceding sources, and the method integrates easily into research models commonly found in entrepreneurship scholarship.

4 Remedies Using Two Stage Least Squares

There are numerous estimators available to deal with endogeneity. The choice of a particular estimator is a function of the endogeneity problem, or problems, faced by the researcher and the study’s research design; for example, bivariate probit models can be particularly effective in dealing with endogeneity in limited dependent variable models (Plummer et al., 2016). We chose to focus though on IV/2SLS for three reasons. As Angrist and Pischke (2008) noted, IV/2SLS methods are flexible enough to address endogeneity in the majority of applied social science analyses. To illustrate, we will demonstrate using IV/2SLS in a direct-effect, latent variable, and mediation model, and discuss its applicability in a moderation model. IV/2SLS also deals with multiple endogeneity threats simultaneously. For example, where the popular Heckman correction addresses sample-selection problems (Shaver, 1998), it does not address self-selection concerns, or other possible sources of endogeneity (Clougherty et al., 2016). In comparison, a well specified 2SLS estimator can generally correct for both types of selection effects, along with other endogeneity threats (Certo et al., 2016).

The second reason we chose to focus on IV/2SLS is because of its prevalence in the literature (Semadeni et al., 2014). IV/2SLS is the workhorse of modern econometrics, and there is a plethora of academic and applied publications discussing how to use the method, its assumptions, software tools for estimating models, and evaluating model robustness. The third reason we chose to focus on IV/2SLS is because, in our opinion, it is a ‘gateway method.’ If a researcher understands the IV/2SLS method and its assumptions, the logic extends easily to more complicated methodologies. For example, Generalized Methods of Moments (GMM)

estimators, of which IV/2SLS can be considered a special case, provides more flexibility in addressing complicated identification problems (Hansen, 1982). The flexibility comes at the expense of complexity, and properly specifying a GMM model requires additional assumptions and more technical knowledge of the estimator (Wooldridge, 2010). Understanding IV/2SLS forms a solid foundation for researchers newer to addressing endogeneity to enhance their methodological toolkit and to improve the causal inference of their research designs.

4.1 Unpacking Two-Stage Least Squares

4.1.1 The 2SLS process

The 2SLS technique gets its name from the series of two equations used in the process:

$$x = \alpha + \beta_{i1} + \beta_{i...} + \Gamma + \epsilon \quad (9)$$

$$y = \alpha + \beta_{\hat{x}} + \Gamma + \epsilon \quad (10)$$

In Eq. 9, we regress our focal endogenous variable, x , on a set of exogenous instrument variables, i , and a vector of exogenous covariates, Γ . We next obtain the predicted values, $\hat{x}$, from Eq. 9, which represents the variance in x due to the exogenous covariates (that must appear in both equations), and the variance in x due to the instruments, which are only used in Eq. 9. We then use $\hat{x}$ in place of x in Eq. 10, and estimate our focal research model accordingly (Bascle, 2008). The 2SLS process, assuming the instruments are valid (relate strongly to x) and are properly excluded (do not correlate with ϵ in the second stage equation), produce consistent estimates of the effect of x on y . The stronger the relationship between the instruments and x the better, because assuming the instruments are exogenous, the more precise, and hence less noisy, the predicted value of x will be in the second stage. While it is possible to perform 2SLS manually, researchers should use specific 2SLS estimators because using $\hat{x}$ in place of x in the second stage biases the standard errors for $\beta_{\hat{x}}$, necessitating a correction.

4.1.2 How instruments work.

Figure 2: The 2SLS Estimation Process

Consider a model with y predicted by x and z , shown in Panel 1 of Figure 2. If we omit z from our model, the estimate for βx will be inconsistent. Omitting z means that the change in y is now a function of change in x plus the unobserved change in z now in the disturbance term of the equation, which we can depict in terms of βx :

$$\beta x = \frac{\Delta y_x + \Delta y_\epsilon}{\Delta x} \quad (11)$$

Assuming that we cannot observe z , the question then becomes whether we can isolate the change in x such that the resulting change in y is due only to variance in x and no other source of variance. The answer is yes, using an instrument variable(s) for x . If the researcher can make x change without any corresponding change in z , then the resulting change in y will be due to the change in x and x alone. We isolate the causal variance in x to predict y by removing the variance in x that is shared with z . Depicted in Panel 2 of Figure 2, here we show two instrument variables, i_1 and i_2 , that causally relate to x , but have no relationship with z . When the instruments cause a change in x , they do so without any corresponding change in z — x changed independently of z . Now when we estimate the relationship between x and y , we have a consistent estimate of βx , because the exogenous change in x results in $\Delta y_\epsilon = 0$.

$$\beta x = \frac{\Delta y_x}{\Delta x} \quad (12)$$

Intuitively, this means that for instruments (i) to be valid, they must strongly relate to x but then have no relationship with z , and by extension, the disturbance term, ϵ , in the final model equation. Importantly, this assumption has little to do with the instrument's covariance with y . It is quite common to observe a zero-order correlation between an instrument and y given the absence of x in a pairwise correlation. The critical question is not whether i and y

correlate, but whether the effect of i on y transmits fully through x . To put it differently, x must **fully** mediate the relationship between i and y such that $i -> x -> y$. If x fills this mediator role, we can think of i as a theoretically valid instrument for x in the context of x 's relationship to y .² Establishing this causal chain depends on a number of assumptions, including the strength of the instruments as predictors of x , which is an issue we turn to next and that carries important implications for entrepreneurship researchers using instrument variable methods.

4.2 Identifying Potential Instrument Variables

There is no easy, simple, or straightforward method to identify valid instruments (Angrist and Pischke, 2008). It is likely to be the most frustrating element of the research design and analysis process, particularly if after collection, the instruments are not valid by random bad luck! Identifying potential instruments is, however, the most important part of the process, and must be a consideration at the earliest stages of the research design process (Semadeni et al., 2014). In other words, when planning an experiment (with an element, such as a mediator, that is not manipulated), a survey, collecting secondary data, or purchasing a dataset, the question of ‘what instruments will I likely need’ and ‘where will I find those instruments’ should be asked early, and asked often.

Recall that for an instrument, i , to be valid, it must strongly relate to the potential endogenous variable, x , but then have no relationship with the unobserved omitted variable, z . Unfortunately we only observe i and x , and so we must depend on a clearly articulated theoretical rationale for why i predicts x , and why i only influences the ultimate outcome variable, y , through its relationship with x : that is, $i -> x -> y$. I do not recommend using the rule of thumb that an instrument must correlate with x but not with y to be valid. While there is some truth to this statement, in practice it can be deceiving, because zero-order correlations are themselves subject to endogeneity (Antonakis et al., 2010). It is very possible, and for some models beneficial, to observe a statistically significant zero-order correlation between i and y , and yet i could be a valid instrument when modeled properly ($i -> x -> y$).

A better approach is to think about potential instruments through the lens of the causal mediation model discussed previously $i -> x -> y$. What would be the variable or variables (i 's) that are likely to cause variation in x , but then any resulting variation in y is from x alone. The pool of valid instruments then is very much a function of the focal relationship between x and y , and the proposed research design to test that relationship. In an experimental design on individual entrepreneurs, potential instruments for x could be intelligence, personality, or perhaps even age (Jacquart and Antonakis, 2015). In a firm-level survey with latent variables, possible instruments for x could be psychometric indicators of a related latent construct to x (Eshima and Anderson, 2017). For secondary, firm-level data, potential instruments for x

²Note that a researcher does not need to actually test for full mediation, because by definition, if the instruments are properly excluded, there would be no ‘direct effect’ present between the instrument and the final outcome variable. If the researcher did test for full mediation, and a direct effect remained, that would indicate that an instrument(s) is not properly excluded and would not be suitable.

could be regulatory or policy changes (Bandiera et al., 2011). The material point is to place the instrument(s) at the start of the causal chain from $i \rightarrow x \rightarrow y$, where x fully mediates the indirect effect of i on y .

The 2SLS method requires at least two instruments per endogenous variable (Angrist and Pischke, 2008). Straight instrument variable estimation requires only one instrument per endogenous variable to identify the model, but a just-identified model does not have the ability to statistically evaluate whether the instruments are properly excluded from the second stage equation. That is, a just-identified model cannot evaluate whether the instrument is itself endogenous, hence the recommendation for two instruments per endogenous variable (Semadeni et al., 2014).

4.3 Evaluating 2SLS Models – Direct Effect with Observed Variables

Figure 3: Common Research Models Integrating 2SLS

I start with evaluating a simple direct effect model, shown in Panel 1 of Figure 3. For each model, I provide simulated data and code in R and in Stata, available online at

<https://osf.io/d453n/>. My purpose here is to describe the basic requirements for instrument validity, to evaluate whether the instruments are properly excluded, and finally to test whether a 2SLS model is necessary (whether the focal predictor is endogenous). I set the number of observations to 1,000 for each simulation.

For this example I used the ‘cross_sectional_sim_data.csv’ dataset, to estimate a direct effect of x on y . This dataset has an omitted variable, z , which is the ‘true’ cause of y and that correlates with x . There are also two instruments (i_1 and i_2) for the 2SLS estimator.

Table 1: 2SLS – Direct Effect

	Model 1 - True Model	Model 2 - Omitted z	Model 3 - 2SLS
Model Variables			
x	-0.014 se = 0.027; p = 0.616	0.151 se = 0.028; p < 0.001	-0.029 se = 0.066; p = 0.661
z	0.526 se = 0.034; p < 0.001		
$c1$	0.094 se = 0.03; p = 0.002	0.082 se = 0.034; p = 0.015	0.08 se = 0.034; p = 0.02
$c2$	0.214 se = 0.031; p < 0.001	0.179 se = 0.034; p < 0.001	0.177 se = 0.035; p < 0.001
$c3$	0.301 se = 0.031; p < 0.001	0.281 se = 0.035; p < 0.001	0.279 se = 0.036; p < 0.001
(Intercept)	0.035 se = 0.031; p = 0.263	0.049 se = 0.034; p = 0.154	0.046 se = 0.035; p = 0.188
First Stage Instruments			
I_1			0.402 se = 0.034; p < 0.001
I_2			0.356 se = 0.037; p < 0.001
Model Fit			
R-squared	0.28	0.111	0.074
Observations	1000	1000	1000
2SLS Diagnostics			
Weak Instruments Test (~F Test)			111.89; p < 0.001
Sargan Test			0.74; p = 0.39
Durbin-Wu-Hausman (Hausman) Test			9.59; p = 0.002

I present the results for our models in Table 1, with Model 1 being the ‘true’ model with the omitted variable, z , included; Model 2 with z excluded; and Model 3 the 2SLS estimate. The results are from analyses conducted with the **AER** package in R (Kleibers and Zeileis, 2008), although Stata’s **ivregress** command yields indistinguishable results and is available online. Model 1 shows the ‘true’ estimate of βx to be effectively zero ($\beta x = -.014$; p = .616), which is consistent with our simulation of no statistically significant effect of x on y . We observe endogeneity in action in Model 2, where omitting z causes βx to increase to .151, which is statistically significant at p < .001. Employing our instruments in the 2SLS in Model 3 yields -.029 for βx , which is statistically indistinguishable from zero (p = .661) and substantively close to the ‘true’ value for βx . Note, however, the larger standard error in the 2SLS model (s.e. = .066). Larger standard errors result from the information loss in the variance of the endogenous variable, and represent the ‘cost’ of using the 2SLS estimator (Angrist and Imbens, 1995). The estimated coefficient estimate will be consistent, but noisier, and hence the higher likelihood of a Type II error.

The diagnostic information for Model 3 is the most important part of Table 1. There are three diagnostic statistics critical to calculate and to interpret in a 2SLS model. Researchers should include this information in a paper, and reviewers and editors should insist on their presence. I’ve summarized each diagnostic statistic and their interpretation in Table 2, and briefly review each below.

Table 2: 2SLS – Diagnostic Statistics

Diagnostic	Description	Rejecting the null means
Weak Instruments Test (~F Test)	Test in the first stage of 2SLS of the joint validity of the instruments, under the null that the two instruments predict x better than an intercept-only model	The instruments exhibit joint validity; however the test statistic should be > 10.0
Sargan (Sargan-Hansen) Test	Test of overidentifying restrictions, under the null hypothesis that the instruments are properly excluded from the second stage of 2SLS (they are endogenous)	The instruments are themselves endogenous (one or more causally predicts y), and are NOT valid
Durbin-Wu-Hausman (Hausman) Test	Test of the consistency of parameter estimates across models, comparing a 2SLS to an OLS model, under the null hypothesis that there is no difference between the two estimators	The 2SLS model statistically differs from OLS, and should be retained; one or more estimates from the OLS model are inconsistent

4.3.1 Weak instruments (Instrument validity)

To be valid, instruments must be individually and jointly significant. We assess individual significance as we would in any regression model—whether the coefficient estimate for each instrument differs significantly from zero. There is no ‘rule of thumb’ here, except to say that the stronger the instrument the better (e.g., $p < .001$). In our model, each instrument is a strong and significant predictor of x ($\beta_{i1} = .402$; $p < .001$; $\beta_{i2} = .356$; $p < .001$). We assess joint significance using an F test of the joint null hypothesis that an intercept only model, expressed as $H_0 : \beta_{i1} = 0$ and $\beta_{i2} = 0$, fits equally well as the model with the instrument-predictors (Murray, 2006). A common rule-of-thumb is an F value greater than 10.0 (Semadeni et al., 2014), and while often appropriate, Stock and Yogo (2005) provide a useful reference determine the specific threshold for the model at hand. Again, larger is better, and in our model, we report an F value of 111.89, which is strongly significant ($p < .001$).

4.3.2 Sargan test of overidentifying restrictions

Assuming the instruments are individually and jointly valid, the next step tests the assumption that the instruments are properly excluded from the second stage equation. The instruments cannot themselves be endogenous, and we use the Sargan test (also called the ‘J’ test or Sargan-Hansen test for related models) under the null hypothesis that the instruments do not materially covary with the disturbance term in the second stage equation. The Sargan test statistic follows a Chi-square distribution with $m - k$ degrees of freedom, where m is the number of instruments and k the number of endogenous variables (Angrist and Pischke, 2008). When m equals k , we cannot conduct the difference test, hence the need for at least two instruments per endogenous variable.

Failing to reject the null hypothesis (e.g., $p > .1$) indicates that there was little statistical evidence that the instruments correlate with the model disturbance term. In this case, the *smaller* the Sargan statistic the better for the model, in terms of instrument validity. Rejecting the null hypothesis (e.g., $p < .05$) suggests that one or more of the instruments are *not* properly excluded and should be carefully reexamined. A rejection of the null hypothesis of the Sargan test also means, in practice, to largely ignore the coefficient estimate of the endogenous variable (Angrist et al., 1996). In our model, we report a Sargan test statistic of

.74 ($p = .39$), suggesting that our instruments are properly excluded and that we meet the assumptions necessary to interpret the coefficient estimate of βx as consistent.

4.3.3 Durbin-Wu-Hausman (Hausman) test for endogeneity

Our last step is to evaluate the assumption that x is endogenous. This test is important because if x is not endogenous, then a more efficient estimator is the better choice because of the information lost in 2SLS estimation (Semadeni et al., 2014). In other words, if the researcher does not have to use an instrument variable method, it is better to use a simpler estimation approach (e.g., OLS). The only way to make this determination, however, is to collect instruments during the research design phase, evaluate the instrument's individual and joint significance, and then specify an instrument variable/2SLS model first, before formally testing the hypothesis. The researcher would also then provide all of these analyses in the paper, including the results of the three test statistics mentioned specifically. *Post hoc* analyses of whether a focal variable is endogenous incorrectly assumes that the variable is exogenous *a priori*; the researcher is effectively attempting to prove the null hypothesis (Antonakis et al., 2010). As Semadeni et al (2014) observed, researchers should start by assuming endogeneity, and then evaluate the evidence to determine whether to relax this assumption.

The Hausman test simply compares the coefficient estimates of the 2SLS model with those from a model using OLS. Often, 2SLS software estimates the OLS model behind-the-scenes, although the researcher could do the procedure manually. The Hausman test makes the assumption that the 2SLS coefficient estimates are consistent, which is why evaluating and reporting the validity of the instruments is crucial *before* employing the Hausman test. The test evaluates the null hypothesis that there is no statistically significant difference between the 2SLS and OLS estimates. Rejecting the null hypothesis (e.g., $p < .05$) suggests that there is a statistically significant difference between the models and that the OLS estimates are inconsistent; the researcher should retain the 2SLS model. Failure to reject the null (e.g., $p > .1$) suggests that the focal variable may not be materially endogenous, and a more efficient estimator (OLS) may be a preferred approach.

In our model, we report a Durbin-Wu-Hausman statistic of 9.59 ($p = .002$), which we expected given the simulated omitted variable, z , that we excluded from the 2SLS model. In our case, the Hausman test provides statistical evidence that x is endogenous to y , and that we should retain our 2SLS model to interpret the effect of x on y . The preceding is the critical point here—a rejection of the null provides statistical evidence that there is an endogenous relationship between x and y . In this case, to evaluate the hypothesis that x impacts y , the researcher should retain the instrument variable/2SLS estimator. Note that it is possible to conduct a Hausman test with a just-identified model (one instrument per endogenous variable). Because, however, evaluating instrument validity with just one instrument is limited to a theoretical argument alone, interpreting a Hausman test if only one instrument is available rests heavily on the theoretical validity of the instrument.

4.4 Latent Variable Model – Direct Effect

Next we evaluate a latent variable model, which estimates a structural (direct) effect of x on y . We set x and y as reflectively measured latent constructs with three indicators each, as shown in Panel 2 of Figure 3. We used the simulated data ‘latent_sim_data.csv’, and estimated our models using the `lavaan` package in R and the `SEM` module in Stata. We chose a latent variable model because many variables in entrepreneurship research are latent in nature, and often measured using psychometric indicators. As such, latent variables have the distinct possibility of being subject to endogeneity from both measurement error and from omitted variable bias. Notably a structural equation modeling (SEM) estimator, like I use here, effectively addresses the former by modeling measurement error, while easily incorporating instrument variables to address the later. It is for this reason why I strongly recommend covariance-based SEM when working with latent constructs, and particularly when working with latent constructs that also suffer from potentially endogenous relationships. The dataset includes an omitted variable, z , and two instrument variables for our latent x : i_1 and i_2 .

Table 3: 2SLS – Direct Effect

	Model 1 - True Model	Model 2 - Omitted z	Model 3 - 2SLS	Model 4 - IV Only
Structural				
$x \rightarrow y$	-0.041 se = 0.032; p = 0.203	0.103 se = 0.032; p = 0.001	-0.094 se = 0.08; p = 0.239	0.098 se = 0.033; p = 0.003
$z \rightarrow y$	0.439 se = 0.038; p < 0.001			
Instruments				
$i1 \rightarrow x$			0.276 se = 0.039; p < 0.001	0.272 se = 0.039; p < 0.001
$i2 \rightarrow x$			0.337 se = 0.041; p < 0.001	0.339 se = 0.041; p < 0.001
Measurement				
$x \rightarrow x_1$	1.0 se = 0	1.0 se = 0	1.0 se = 0	1.0 se = 0
$x \rightarrow x_2$	0.894 se = 0.045; p < 0.001	0.893 se = 0.046; p < 0.001	0.917 se = 0.046; p < 0.001	0.914 se = 0.046; p < 0.001
$x \rightarrow x_3$	0.908 se = 0.045; p < 0.001	0.928 se = 0.048; p < 0.001	0.936 se = 0.047; p < 0.001	0.935 se = 0.047; p < 0.001
$y \rightarrow y_1$	1.0 se = 0	1.0 se = 0	1.0 se = 0	1.0 se = 0
$y \rightarrow y_2$	1.082 se = 0.066; p < 0.001	1.065 se = 0.07; p < 0.001	1.058 se = 0.069; p < 0.001	1.066 se = 0.07; p < 0.001
$y \rightarrow y_3$	0.96 se = 0.061; p < 0.001	0.955 se = 0.063; p < 0.001	0.953 se = 0.063; p < 0.001	0.955 se = 0.063; p < 0.001
Covariances				
$z \sim x$	0.484 se = 0.048; p < 0.001			
$y \sim x$			0.287 se = 0.109; p = 0.009	
Model Fit				
Observations	1000	1000	1000	1000
Model Chi-sq	14.536 df = 12; p = 0.268	7.808 df = 8; p = 0.452	18.584 df = 17; p = 0.353	25.947 df = 18; p = 0.101

I present the results for our latent variable direct effect model in Table 3, with Model 1 being the ‘true’ model with the omitted variable, z , included; Model 2 with z excluded; and Model 3 the 2SLS estimate. I set the scale of the latent variables in each model by setting a loading for one indicator to 1.0 (Bollen, 1989). Note specifically the inclusion of the covariance between the disturbance term for x and for y in the 2SLS model. Estimating this parameter is critical—without it, the model assumes the covariance is zero between x and the disturbance term for y (that x is strictly exogenous to y). Not freeing the parameter, then, is the same thing as estimating a naive model, *even if* instruments are included as predictors of x . In short, this parameter must be freely estimated to be a valid 2SLS model. Model 4 shows an example of omitting this covariance, which yields an estimate for βx substantially similar to the omitted variable Model 2, despite including instruments in the model.

As with our earlier example, the ‘true’ model with z included yields a non-significant estimate for βx ($-.041$; $p = .203$). With z omitted, Model 2 inflates the estimate for βx and flips signs to $.103$, which is statistically significant ($p < .001$). Lastly, our 2SLS estimate in Model 3 yields an estimate closer to our true model of $-.094$ ($p = .239$). We again rely on our three diagnostic statistics to evaluate the 2SLS latent variable model. Because covariance-based SEM employs maximum likelihood (ML) for estimation, we describe the ML counterparts to the least square statistics discussed previously, and summarize these below and in Table 4.

Table 4: 2SLS – SEM/Maximum Likelihood Diagnostic Statistics

SEM statistic	Which is equivalent to the...	Model value
Wald Chi-sq	Weak Instruments Test ($\sim F$ Test)	113.77 df = 2; $p < 0.001$
Model Chi-sq	Sargan Test	18.584 df = 17; $p = 0.353$
Chi-sq Difference	Durbin-Wu-Hausman (Hausman) Test	7.363 df = 1; $p = 0.007$

4.4.1 Evaluating instrument validity

As before, instruments must be individually and jointly significant to be valid. As reported in Model 3 of Table 3, our two instruments are strong, significant predictors of the latent variable, x ($\beta_{i1} = .276$; $p < .001$; $\beta_{i2} = .337$; $p < .001$). However, there is no F test in the first stage to evaluate joint instrument validity with a ML estimator. Maximum likelihood does provide a cousin to the F test, the Wald test, that is an appropriate substitute for the F test in SEM models (Bollen, 1989). A very flexible test statistic, we can use the Wald statistic to test the joint significance of the coefficient estimate for our instruments against an intercept-only model, which is conceptually similar to the first stage F test in the least squares 2SLS estimator.

An F statistic multiplied by the test degrees of freedom is asymptotically equivalent to a Chi-squared statistic at the equivalent degrees of freedom (Cohen et al., 2003). With two instruments in our model ($df = 2$), the Wald Chi-square statistic is roughly equivalent to two times the F value, or $F = \frac{\chi^2}{df}$. Drawing from our diagnostic results in Table 4, the preceding formula yields an F statistic of 56.885, which would be well above the recommended critical threshold of 10.0.

4.4.2 Overidentifying restrictions (Sargan test)

As earlier, assuming the instruments are individually and jointly valid, the next step is to determine whether each instrument is properly excluded (i.e., whether the instruments are endogenous). For SEM models, we recommend a two-step approach to evaluate this assumption. In the first step, we examine the global fit of the model using the Chi-squared statistic. The benefit of using the model Chi-square is that if the instruments did covary with the disturbance term for y , the model Chi-square would inflate, indicating a misspecification because the researcher incorrectly assumed that covariance to be equal to zero. In other

words, a significant Chi-square statistic (e.g., $p < .05$) tells the researcher that he/she has made an error in model specification—the researcher’s model is significantly different from the underlying covariance structure of the data. This is one of many reasons why the model Chi-square is so critical to evaluating the global fit of an SEM model, and why I recommend strongly against using normative fit indices (e.g., CFI, RMSEA) to evaluate fit (Antonakis et al., 2014; Williams and O’Boyle, 2011). As reported in Table 4, our model Chi-square indicates no evidence of material misspecification (18.584; $df = 17$; $p = .353$).

The Chi-squared test provides a global perspective—rejecting the null hypothesis (e.g., $p < .05$) tells the researcher a misspecification is present, but not which parameter is misspecified. One approach is to use a series of nested model comparisons, sequentially freeing the covariance between each instrument and the disturbance term for y and testing the resulting model against the original model. My preferred approach, however, is to make use of a tool found in virtually all SEM software, and that is the modification indices. The modification indices systematically evaluate whether freeing a given parameter results in a statistically significant decrease in the model Chi-squared value. If the covariance between the instrument and the disturbance term appears in the modification indices output, it tells the researcher that the instrument is itself endogenous and hence likely not valid. In summary, to “pass” the Sargan test with an SEM model, the researcher would observe a non-significant Chi-square statistic, *and* there would be no covariance between an instrument and the disturbance term for y listed in the model modification indices, which is the case for Model 3.

4.4.3 Testing for endogeneity (Durbin-Wu-Hausman test)

The last step is to evaluate our assumption that x is endogenous to y . As before, if x is not endogenous, then estimating the model without instruments yields a lower standard error and is more efficient. The parameter of interest here is the covariance between the disturbance terms for x and y . If this covariance is statistically significant (e.g., $p < .05$), then the researcher may infer that x is endogenous to y ; if it was not statistically different from zero (e.g., $p > .1$), there would be no need to include the covariance, because x does not materially relate to the disturbance term for y (x likely is exogenous to y). As reported in Model 3 of Table 3, we observe a statistically significant covariance between the two disturbance terms (.287; $p = .009$)—as expected given the omission of z —and we can safely assume that we should retain the 2SLS model.

A more formal way to make this determination is a Chi-square difference test between a model with the parameter freely estimated (Model 3) and a model with the parameter constrained to be zero (Model 4). This test is equivalent to the Hausman test discussed previously, and a statistically significant test statistic (e.g., $p < .05$) provides evidence that the focal covariance statistically differs from zero. As reported in Table 4, our Chi-square difference test confirms our earlier determination that x is endogenous to y with a significant difference between the two models (7.363; $df = 1$; $p = .007$).

4.5 Mediation Model

Lastly I estimate an observed-variable mediation model where the indirect effect of x on y proceeds through m , or equivalently $x \rightarrow m \rightarrow y$, shown in Panel 3 of Figure 3. We used the simulated data ‘mediation_sim_data.csv’, and estimated our models using the `lavaan` package in R and the `SEM` module in Stata. In this dataset, I included an omitted variable, z , which in this analysis represents an omitted mediator between x and y ; z is the ‘true’ causal mechanism connecting x and y . I also include instruments for x : i_1 and i_2 , and for m : i_3 and i_4 .

4.5.1 Model specification

For this mediation model, neither x nor m are experimentally manipulated. We treat them as measured, or observational, as if we collected the data through a survey, or used secondary data. As previously discussed, with any observational data, the researcher never meets the assumption that omitted variables are not biasing the observed relationships. As such, both x and m require instrument variables. If, for example, the researcher manipulated x in an experiment, assuming randomized assignment, it is not necessary to instrument x . If in this same experiment the researcher measured m by, for example, administering a survey, m remains an observed variable and requires instruments. If endogeneity biases the estimate of the relationship between x and m , or between m and y , the resulting indirect (mediating) effect will itself be inconsistent (Imai et al., 2010).

4.5.2 Model results

Table 5 displays the mediation model results, with Model 1 being the ‘true’ model with the omitted mediator, z , included; Model 2 with z excluded; and Model 3 the 2SLS estimate. With the mediation model, we freely estimate all possible disturbance term covariances, because the estimator—whether I included a ‘direct effect’ of x on y or not—assumes that all possible predictor variables do not correlate with each structural disturbance term (Bollen, 1989). As such, we free the possible covariances between the disturbance terms of x and m ; x and y ; and m and y .

The focal parameter of interest is the indirect effect path (ab following the familiar Baron and Kenny (1986) notation), calculated by multiplying the estimated coefficient for the relationship between x and m (the a path), and the relationship between m and y (the b path). We infer a material indirect effect of x on y through m if the ab path is significantly different from zero (Zhao et al., 2010). As shown in Model 1 with the ‘true’ mediator, z , included, the indirect effect effectively equals zero (.000; $p = .81$). With z omitted in Model 2, the indirect effect inflates and is statistically significant (.023; $p = .001$). In our 2SLS estimate in Model 3, the indirect effect is once again close to its true value and not significant (-.001; $p = .713$). We display our 2SLS diagnostic statistics in Table 6, and briefly discuss these values below.

Table 5: 2SLS – Mediation

	Model 1 - True Model	Model 2 - Omitted z	Model 3 - 2SLS
Structural			
x → m	-0.009 se = 0.03; p = 0.754	0.124 se = 0.031; p < 0.001	-0.027 se = 0.066; p = 0.679
m → y	0.01 se = 0.028; p = 0.707	0.185 se = 0.028; p < 0.001	0.049 se = 0.06; p = 0.417
z → m	0.548 se = 0.04; p < 0.001		
z → y	0.556 se = 0.037; p < 0.001		
Instruments			
i1 → x			0.431 se = 0.034; p < 0.001
i2 → x			0.28 se = 0.035; p < 0.001
i3 → m			0.379 se = 0.034; p < 0.001
i4 → m			0.437 se = 0.033; p < 0.001
Indirect Effect			
a*b → ab	0.000 se = 0.000; p = 0.81	0.023 se = 0.007; p = 0.001	-0.001 se = 0.004; p = 0.713
Covariances			
z ∼ x	0.377 se = 0.039; p < 0.001		
x ∼ m			0.228 se = 0.093; p = 0.014
y ∼ m			0.215 se = 0.083; p = 0.009
x ∼ y			0.205 se = 0.042; p < 0.001
Model Fit			
Observations	1000	1000	1000
Model Chi-sq	0.054 df = 1; p = 0.817	18.624 df = 1; p < 0.001	5.094 df = 6; p = 0.532

Table 6: 2SLS – Mediation Diagnostic Statistics

Diagnostic	Statistic
Wald (Weak Instruments) Test	
Instruments for x	212.49 df = 2; p < 0.001
Instruments for m	282.824 df = 2; p < 0.001
Overidentifying Restrictions (Sargan) Test	
Model Chi-sq	5.094 df = 6; p = 0.532
Hausman (Chi-squared Difference) Test	
Constraining x ∼ m = 0	6.125 df = 1; p = 0.013
Constraining m ∼ y = 0	6.842 df = 1; p = 0.009
Constraining x ∼ y = 0	25.099 df = 1; p < 0.001

4.5.3 Evaluating 2SLS diagnostics in mediation

Because our SEM model uses Maximum Likelihood, we interpret our diagnostic statistics in Table 6 similar to the latent variable model. The instruments are individually significant as shown in Model 3 of Table 5, i_1 and i_2 for x at .431 ($p < .001$) and .28 ($p < .001$) respectively; and i_3 and i_4 for m at .379 ($p < .001$) and .437 ($p < .001$) respectively. For joint significance, we estimate two different models with a Wald test evaluating the joint null hypothesis that the pair of instruments for each respective endogenous variable explain more variance in the endogenous variable than would an intercept-only model (Cohen et al., 2003). As shown in Table 6, both Wald tests are large (e.g., > 20) and statistically significant (x : 212.49; $df = 2$;

$p < .001 \mid m: 282.82; df = 2; p < .001$).

We then use the same two-step methodology to evaluate the overidentification restrictions of the instruments. The global Chi-squared statistic for the model indicates no material misspecification (5.094; $df = 6$; $p = .532$). Further, the modification indices display no possible instrument covariance between any potential instrument and all of the equation-level disturbance terms. Taken together, we interpret our results that there is no statistical evidence in the model suggesting that the instruments are themselves endogenous (and invalid).

Lastly, a series of Chi-square difference tests evaluates each possible endogenous path by asking whether the model with freely estimated covariances differs significantly from a model where we constrain each covariance to be zero in turn. This procedure mimics a Hausman test for each covariance. As shown in Table 6, each comparison is statistically significant, as expected, given that we omitted the ‘true’ causal mechanism, z . As such, we must retain the 2SLS estimator to recover a consistent estimate of the indirect effect.

4.5.4 Value of covariance-based structural equation modeling for evaluating mediation

Covariance-based SEM has a number of advantages over the stepwise regression approach outlined by Baron and Kenny (1986), and the bootstrapped standard errors method suggested by Zhao et al. (2010) to test mediation. Further, covariance-based SEM easily incorporates bootstrapping and related methods to estimate robust standard errors. Given the likelihood of the indirect effect being biased by endogeneity in the a path, the b path, or both, we recommend against using estimators (e.g., stepwise regression, the PROCESS macro, PLS) that cannot accommodate instrument variable models. Covariance-based SEM in R (`lavaan`), Stata (`SEM`), SPSS (`AMOS`), and MPlus will do all that the non-SEM estimators will do and more. In short, there is nothing to be gained, but much to be lost, by testing mediation without using covariance-based SEM.

4.6 Moderation Model

A proper treatment of 2SLS in moderation is itself a separate paper, because endogeneity manifests very differently depending on whether one or both of the constituent variables are endogenous, and the degree of that endogeneity (Withers et al., 2014). A related case is the so-called ‘forbidden regression,’ which occurs when an endogenous variable is used as a linear combination of itself, such as in a curvilinear relationship (Wooldridge, 2010). We refer interested readers to Wooldridge (2010) for more information on properly specifying these models, but given the popularity of these models in entrepreneurship research, we offer some best practice recommendations for estimating moderating effects while accounting for endogeneity using IV/2SLS.

Consider the standard equation for an interaction effect:

$$y = \alpha + \beta x + \beta m + \beta xm + \epsilon \quad (13)$$

Here x and m are predictor variables, and xm is the multiplicative interaction term of the two predictors. Following Wooldridge (2010), we suggest evaluating the endogeneity of x and of m as separate endogenous predictors but in a single specified model. In other words, the 2SLS model would have two potentially endogenous predictors and at least four instruments; two for each predictor. The researcher then uses the standard 2SLS diagnostics to determine whether to treat x , m , or both as endogenous. If neither term is endogenous, their product will not be either (Wooldridge, 2010). As such, the researcher would report the results of the 2SLS models, and in the absence of statistical evidence of endogeneity, proceed to test the research model using a non-2SLS estimator (Titus and Anderson, 2016). Similarly for a curvilinear relationship, the researcher would first evaluate the endogeneity of x in a linear model, and if there is no statistical evidence of endogeneity, proceed to testing the curvilinear effect.

At the other extreme, if both x and m are endogenous, it becomes problematic to test this interaction effect. While it is possible to create a vector of instruments to identify this model, the resulting instrumented interaction term takes on very little practical meaning (Angrist and Pischke, 2008). It would not be appropriate to interpret the instrumented interaction coefficient ($\beta_{x\hat{m}}$) as the average change in the in the slope of x on y as a function of a unit change in m , because $\beta_{x\hat{m}}$ is itself only a portion of the total variance in xm (Angrist and Pischke, 2008; Wooldridge, 2010). As such, the results of the model become very difficult, if not largely meaningless, to interpret practically.

There remains some debate about specifying interaction models with one endogenous predictor, and one exogenous predictor (Withers et al., 2014), and here we refer readers to Wooldridge (2010), for further explanation. It is possible to specify these models, generally using additional constituent interaction terms of other excluded instruments for identification (Angrist and Pischke, 2008). Again though, while the resulting instrumented interaction coefficient ($\beta_{x\hat{m}}$) would be consistent in this case, its interpretation depends heavily on whether it is the focal predictor, x , that was endogenous to y , or if it was the moderator, m (Withers et al., 2014). In these situations we encourage researchers to proceed with caution in their interpretation of these results, and to pay careful attention to the estimation of the marginal effects and confidence intervals around the estimated parameters.

4.7 Summary

Importantly, the examples in this paper deal with continuously measured predictors and outcome variables. Models using nominal, dichotomous, or ordinal predictors and outcomes generally require specialty estimators (Angrist, 2001; Angrist and Pischke, 2008). The simulated models are also basic by design. In most cases, however, the underlying logic of 2SLS estimation extends to more complicated IV models (Angrist et al., 1996). However, researchers employing instruments in more complicated research designs (e.g., moderated-mediation) should ensure that the simplest relevant model (e.g., mediation) meets the

necessary assumptions for instrument validity. The best approach is to start with a simple, properly specified model and build from there.

5 Discussion & Recommendations

5.1 Challenges and Remedies

Standards are rapidly increasing for publishing research in top management and entrepreneurship journals (Lindsay, 2017; Meyer et al., 2017). Established practices that were commonplace data science in years past are, in many cases, being questioned (Simmons et al., 2011). The pace at which statistical theory—and statistical software—advance also makes it difficult for applied researchers to keep updating their methodological knowledge and skills, despite a desire to publish high quality, accurate research that makes a meaningful theoretical contribution (Bettis et al., 2016). Reviewers, and editors, often deal with conflicting opinions of best practice, increasing the variance in the review process and making it difficult for authors to meet the quality expectation of an academic journal (Meyer et al., 2017).

One purpose of this paper was to demonstrate the multiple ways endogeneity may influence entrepreneurship research. Particularly concerning is the self-selection effects inherent to our phenomenon—individuals self-select into being entrepreneurs, and firms determine their entrepreneurial orientation. Choosing to ignore selection effects and other sources of endogeneity is no longer an option for entrepreneurship researchers. Incorrect or misspecified models impact the usefulness of a given paper, and also ‘spill-over’ to meta-analyses, literature reviews, and future studies (Bosco et al., 2016). The following recommendations then are ‘best practices’ as it relates to dealing with endogeneity and threats to causal inference in entrepreneurship research.

5.2 Take Endogeneity Seriously

The most important recommendation for editors and reviewers is determining whether the author took endogeneity and omitted variable bias seriously (Semadeni et al., 2014). For authors, the converse holds that we must demonstrate to reviewers and to editors that we have taken every reasonable step to recover consistent parameter estimates. In any observational research, or in an experimental design that includes a measured (not manipulated) variable, endogeneity is possible and likely (Rubin, 2005). Practically speaking, that means that authors should be reporting some type of consideration, at the research design stage, to address endogeneity, be it with instrument variables, a natural experiment, or other related technique. Because the credibility of our field rests, in no small part, on the quality of our science (Antonakis, 2017), meaningfully addressing endogeneity in observation research should be a *de facto* standard. Further, in the case of models with multiple predictors (or mediators/moderators), each hypothesized path should have a corresponding discussion of why—or why not—endogeneity is a concern in that specific relationship.

Reviewers should, however, avoid blanket pronouncements about endogeneity being present in a study. The question of whether or not endogeneity is present is not likely to be helpful to an author, because the answer is generally yes (Antonakis et al., 2010). Rather, reviewers should help the author identify and quantify the specific source or sources of endogeneity for a given research design and estimation approach. This is particularly the case when an author may have addressed one type of endogeneity concern, for example, using a Heckman correction to deal with a selection effect, but may not have addressed other concerns, such as confounding variables. The objective for the reviewer is to help educate the author on how endogeneity specifically threatens the efficacy of his or her research design, and also how to constructively—and correctly—address those concerns.

5.3 Align the Assumptions of the Method to the Research Design

The second recommendation, assuming some method to deal with endogeneity in a study, is to ensure that the assumptions of the method are congruent with the research design and data collected. In the case of instrument variable approaches, including 2SLS, a natural experiment, or regression discontinuity analysis, the critical question is whether the instrument(s) is theoretically expected to induce variance in the endogenous variable, independent of the possible omitted variable(s) (Angrist et al., 1996). That is, there must be a compelling conceptual argument for why the instrument relates to the outcome *only* through the endogenous variable ($i \rightarrow x \rightarrow y$). Absent theoretical justification, the instruments may simply be an artifact of the dataset or design, which limits their usefulness to other researchers and raises the possibility of a spurious relationship from the instrument to the endogenous variable (Angrist and Pischke, 2008).

In practice, this means that researchers must be considering endogeneity remedies, and potential instruments, early in the research design process and *before* beginning data collection. For most surveys and other observational data, it is impractical, if not impossible, to ‘go back in time’ and collect more data for instruments. Thinking through the estimation approach and building a theoretical rationale for potential instruments, also referred to as solving the identification problem (Angrist and Krueger, 2001), should go hand-in-hand with identifying a meaningful gap in the literature and how a study addresses this gap.

5.4 Provide Empirical Justification for the Instruments

The third recommendation is to ensure that a given paper supports the conceptual argument for the selected endogeneity remedy with appropriate empirical evidence. In the case of 2SLS and other instrument variable methods, that means providing the following critical pieces of information: 1) the individual significance of the instruments (the estimated coefficient of the instrument on the endogenous variable) and the joint significance of the instruments (the Weak instrument, F / Wald test); and 2) the test of overidentifying restrictions (the Sargan/Sargan-Hansen test). Absent this information, a reader has no way to evaluate whether the instruments are doing the job they are intended for in the model.

We also recommend authors employ and report the result of a Hausman test when using 2SLS, as we did in this paper. The reason being is to give the researcher flexibility, assuming that there is no statistical evidence of material endogeneity, to employ more efficient estimators. However, the choice of a non-instrument variable estimator is *only* appropriate after a thorough evaluation of endogeneity and reporting those results in the paper. *Post hoc* measures to evaluate endogeneity make the assumption that the naive model is consistent, which is not appropriate (Semadeni et al., 2014).

5.5 Making the Most of IV/2SLS Estimators

While powerful, 2SLS and related estimators have substantive limitations. One of which we discussed previously, namely that 2SLS is a limited information estimator, and the higher estimated standard errors raise the possibility of increased Type II error (Gormley and Matsa, 2014). Another limitation is that instruments simply act as proxies for randomization; instruments are subject to measurement error, bias, and other threats to their validity as a proxy (Rubin, 2005). As such, instrument variable methods preclude establishing direct causal relationships—the best a researcher can hope for with instruments is to demonstrably improve causal inference (Winship and Morgan, 1999). Improved inference is meaningful, however, and represents an important advancement in the accuracy of entrepreneurship research.

The preceding limitations mean that making practical use of instruments is complicated. The assumptions underlying recovering consistent parameter estimates are easily violated, and valid instruments in one model may not be valid in the next because of random variation (Angrist and Pischke, 2008). This is why it is valuable for authors to start with simple models in their theory construction before pursuing more advanced specifications. Consider, for example, a conditional indirect effect (i.e., moderated-mediation) model with observational data. In this model, x may be endogenous to y ; m (mediator) endogenous to y ; x endogenous to m ; the moderator, w , endogenous to m ; and w endogenous to y . That’s a lot of potential endogeneity to deal with.

One way to tackle the preceding model would be to take each path in isolation, identify appropriate instruments for that path, and then ensure the instruments are valid and properly excluded. Once properly specified, the researcher proceeds in a stepwise fashion, gradually adding model complexity until reaching the fully specified model, with all relationships tested simultaneously. This is a non-trivial exercise, but one that we feel will make the endogeneity problem more manageable for researchers, and more communicable to reviewers and editors. Given the space limitations in most journal articles, making use of data repositories, such as the *Open Science Framework* used in this paper, to share supporting materials and analyses makes it easier to provide additional detail about an analytical approach.

5.6 Conclusion

My purpose with this paper was to discuss common endogeneity problems and the ways to address it in entrepreneurship research using an IV/2SLS approach. Dealing with endogeneity has become—and rightly so—a *de facto* standard for publishing in top journals, and instrument variable methods are one tool applicable across many common entrepreneurship research models. I certainly encourage scholars to go beyond the 2SLS method outlined here and explore the multitude of estimation options to address endogeneity and to improve causal inference, but I hope that this paper is a useful guide for entrepreneurship scholars in evaluating and dealing with endogeneity in their research and in the research they review.

6 References

- Anderson, B.S., Eshima, Y., 2013. The influence of firm age and intangible resources on the relationship between entrepreneurial orientation and firm growth among Japanese SMEs. *J. Bus. Venturing* 28, 413–429.
- Angrist, J.D., 2001. Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. *J. Bus. Econ. Stat.* 19, 2–28.
- Angrist, J.D., Imbens, G.W., 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *J. Am. Stat. Assoc.* 90, 431–442.
- Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* 91, 444–455.
- Angrist, J., Krueger, A.B., 2001. Instrumental variables and the search for identification: From supply and demand to natural experiments. *J. Econ. Perspect.* 15, 69–85.
- Angrist, J., Pischke, J.-S., 2008. *Mostly harmless econometrics: an empiricist’s companion*. Princeton University Press, Princeton, NJ.
- Antonakis, J., 2017. On doing better science: From thrill of discovery to policy implications. *Leadersh. Q.* 28, 5–21.
- Antonakis, J., Bendahan, S., Jacquart, P., Lalive, R., 2010. On making causal claims: A review and recommendations. *Leadersh. Q.* 21, 1086–1120.
- Antonakis, J., Bendahan, S., Jacquart, P., Lalive, R., Day, D.V., 2014. Causality and endogeneity: Problems and solutions, in: Day, D.V. (Ed.), *The Oxford Handbook of Leadership and Organizations*. Oxford University Press, New York, pp. 93–117.
- Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econ. Stud.* 58, 277–297.
- Atinc, G., Simmering, M.J., Kroll, M.J., 2012. Control variable use and reporting in macro and micro management research. *Organizational Research Methods* 15, 57–74.
- Bandiera, O., Barankay, I., Rasul, I., 2011. Field experiments with firms. *J. Econ. Perspect.* 25, 63–82.
- Baron, R.M., Kenny, D.A., 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51, 1173–1182.
- Bascle, G., 2008. Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization* 6, 285–327.
- Bergh, D.D., Sharp, B.M., Aguinis, H., Li, M., 2017. Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization* Forthcoming, 1–21.

- Bettis, R.A., 2012. The search for asterisks: Compromised statistical tests and flawed theories. *Strat. Mgmt. J.* 33, 108–113.
- Bettis, R.A., Ethiraj, S., Gambardella, A., Helfat, C., Mitchell, W., 2016. Creating repeatable cumulative knowledge in strategic management. *Strat. Mgmt. J.* 37, 257–261.
- Bollen, K.A., 1989. *Structural Equations with Latent Variables*. John Wiley, New York.
- Bosco, F.A., Aguinis, H., Field, J.G., Pierce, C.A., Dalton, D.R., 2016. HARKing's threat to organizational research: Evidence from primary and meta-analytic sources. *Pers. Psychol.* 69, 709–750.
- Boyd, B.K., Bergh, D.D., Ireland, R.D., David J. Ketchen, J., 2013. Constructs in strategic management. *Organizational Research Methods* 16, 3–14.
- Butler, N., Delaney, H., Spoelstra, S., 2017. The gray zone: Questionable research practices in the business school. *Acad. Manage. Learning & Education* 16, 94–109.
- Certo, S.T., Busenbark, J.R., Woo, H.-S., Semadeni, M., 2016. Sample selection bias and Heckman models in strategic management research. *Strat. Mgmt. J.* 37, 2639–2657.
- Certo, S.T., Withers, M.C., Semadeni, M., 2017. A tale of two effects: Using longitudinal data to compare within- and between-firm effects. *Strat. Mgmt. J.* 38, 1536–1556.
- Clougherty, J.A., Duso, T., Muck, J., 2016. Correcting for self-selection based endogeneity in management research: Review, recommendations and simulations. *Organizational Research Methods* 19, 286–347.
- Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Erlbaum, Mahwah, NJ.
- Cook, T.D., Campbell, D.T., 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin, Boston, MA.
- Eshima, Y., Anderson, B.S., 2017. Firm growth, adaptive capability, and entrepreneurial orientation. *Strat. Mgmt. J.* 38, 770–779.
- Goldfarb, B., King, A.A., 2016. Scientific apophenia in strategic management research: Significance tests & mistaken inference. *Strat. Mgmt. J.* 37, 167–176.
- Gormley, T.A., Matsa, D.A., 2014. Common errors: How to (and not to) control for unobserved heterogeneity. *Rev. Financ. Stud.* 27, 617–661.
- Hansen, L.P., 1982. Large sample properties of Generalized Method of Moments estimators. *Econometrica* 50, 1029–1054.
- Holland, P.W., 1986. Statistics and causal inference. *J. Am. Stat. Assoc.* 81, 945–960.
- Imai, K., Keele, L., Tingley, D., 2010. A general approach to causal mediation analysis. *Psychol. Methods* 15, 309–334.
- Imai, K., King, G., Stuart, E.A., 2008. Misunderstandings between experimentalists and observationalists about causal inference. *J. R. Stat. Soc. Ser. A Stat. Soc.* 171, 481–502.

- Ioannidis, J.P.A., 2016. Why most clinical research is not useful. *PLoS Med* 13, e1002049.
- Jacquart, P., Antonakis, J., 2015. When does charisma matter for top-level leaders? Effect of attributional ambiguity. *Acad. Manage. J.* 58, 1051–1074.
- Kleiber, C., Zeileis, A., 2008. *Applied Econometrics with R*. Springer-Verlag, New York, NY.
- Krueger, N., 1993. The impact of prior entrepreneurial exposure on perceptions of new venture feasibility and desirability. *Entrepreneurship: Theory and Practice* 18, 5–22.
- Lindsay, D.S., 2017. Sharing data and materials in psychological science. *Psychol. Sci.* Forthcoming, 1–4.
- MacKenzie, S.B., Podsakoff, P.M., Podsakoff, N.P., 2011. Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly* 35, 293–334.
- Martin, B.C., McNally, J.J., Kay, M.J., 2013. Examining the formation of human capital in entrepreneurship: A meta-analysis of entrepreneurship education outcomes. *J. Bus. Venturing* 28, 211–224.
- Meyer, K.E., van Witteloostuijn, A., Beugelsdijk, S., 2017. What’s in a p? Reassessing best practices for conducting and reporting hypothesis-testing research. *J Int Bus Stud* Forthcoming, 1–17.
- Murray, M.P., 2006. Avoiding invalid instruments and coping with weak instruments. *J. Econ. Perspect.* 20, 111–132.
- Nuijten, M.B., Hartgerink, C.H.J., van Assen, M.A.L.M., Epskamp, S., Wicherts, J.M., 2016. The prevalence of statistical reporting errors in psychology (1985–2013). *Behav Res* 48, 1205–1226.
- Pearl, J., 2014. Interpretation and identification of causal mediation. *Psychol. Methods* 19, 459–481.
- Plummer, L.A., Allison, T.H., Connelly, B.L., 2016. Better together? Signaling interactions in new venture pursuit of initial external capital. *Acad. Manage. J.* 59, 1585–1604.
- Podsakoff, P.M., MacKenzie, S.B., Podsakoff, N.P., Lee, J.Y., 2003. The mismeasure of man(agement) and its implications for leadership research. *Leadersh. Q.* 14, 615–656.
- Roundy, P.T., Brockman, B.K., Bradshaw, M., 2017. The resilience of entrepreneurial ecosystems. *Journal of Business Venturing Insights* 8, 99–104.
- Rubin, D.B., 2005. Causal inference using potential outcomes. *J. Am. Stat. Assoc.* 100, 322–331.
- Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* 66, 688.
- Semadeni, M., Withers, M.C., Certo, T.S., 2014. The perils of endogeneity and instrumental variables in strategy research: Understanding through simulations. *Strat. Mgmt. J.* 35, 1070–1079.

- Shaver, J.M., 1998. Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival? *Manage. Sci.* 44, 571–585.
- Shepherd, D.A., 2015. Party on! A call for entrepreneurship research that is more interactive, activity based, cognitively hot, compassionate, and prosocial. *J. Bus. Venturing* 30, 489–507.
- Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22, 1359–1366.
- Stock, J., Yogo, M., 2005. Testing for weak instruments in linear IV regression, in: Andrews, D. (Ed.), *Identification and Inference for Econometric Models*. Cambridge University Press, New York, pp. 80–108.
- Titus, V.K., Anderson, B.S., 2016. Firm structure and environment as contingencies to the corporate venture capital–parent firm value relationship. *Entrepreneurship Theory and Practice* Forthcoming, 1–25.
- Wiklund, J., Shepherd, D.A., 2011. Where to from here? EO-as-experimentation, failure, and distribution of outcomes. *Entrepreneurship Theory and Practice* 35, 925–946.
- Williams, L.J., O’Boyle, E., Jr, 2011. The myth of global fit indices and alternatives for assessing latent variable relations. *Organizational Research Methods* 14, 350–369.
- Winship, C., Morgan, S.L., 1999. The estimation of causal effects from observational data. *Annu. Rev. Sociol.* 25, 659–706.
- Withers, M., Certo, S.T., Semadeni, M., 2014. Examining the influence of endogeneity when testing interactions. Presented at the Strategic Management Society International Conference.
- Wooldridge, J.M., 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.
- Zhao, H., Seibert, S.E., Hills, G.E., 2005. The mediating role of self-efficacy in the development of entrepreneurial intentions. *J. Appl. Psychol.* 90, 1265–1272.
- Zhao, X., Lynch, J., Chen, Q., 2010. Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *J. Consum. Res.* 37, 197–206.