



EDITORIAL

From the Editors: Endogeneity in international business research

David Reeb,
Mariko Sakakibara and
Ishtiaq P Mahmood

Area Editors

Correspondence:

D Reeb, NUS Business School, National University of Singapore, 15 Kent Ridge Drive, BIZ 1 #7-2, Mochtar Riady Building, Singapore 119245.

Tel: + 65 6516 3042;

Fax: + 65 6779 2083;

Email: bizdmr@nus.edu.sg

Abstract

This essay builds on the exposition by Thomas et al. and focuses on analyzing cause and effect in international business research. We attempt to explain how endogeneity problems occur and why they are so prevalent in international business research in a non-technical fashion. We then discuss the importance of explicitly identifying how the chosen research design best approximates a randomized-controlled experiment. Finally, we provide some guidelines on achieving this goal and emphasize the practices that seem most relevant to *JIBS* reviewers in evaluating high-quality international business research.

Journal of International Business Studies (2012) 43, 211–218. doi:10.1057/jibs.2011.60

Keywords: international business research; research design; endogeneity

THE IDEAL RESEARCH SETTING

Empirical research in international business (IB) is difficult. Our interests typically center on whether some particular IB phenomenon causes a specific outcome or effect. We might, for instance, be interested in how expatriate postings influence future career opportunities. Or we might be seeking to understand how firm-level internationalization affects corporate decision-making. In an ideal research setting, to test such a cause and effect, we would examine the impact of firm internationalization on a particular outcome (such as profitability) by randomly assigning some firms to be multinational corporations (MNCs) and other firms to be domestic corporations (DCs). Experimentalists would characterize these as the treatment and control groups. Preferably, we would then observe and compare the subsequent decision-making of the firms in the treatment and control groups over the next few years regarding the specific variable of interest (i.e., profitability). Inherent in this approach is the notion that we would randomly select the firms to place into the treatment and control groups (i.e., the MNC and DC groupings) in our sample. In general, the iconic test procedure involves developing a random experiment, regardless of whether the unit of analysis centers on individuals, firms, industries or countries.

Unfortunately, in international business research, we are seldom afforded the luxury of a randomized controlled experiment. In addition, in many business situations the treatment may not be a simple binary choice – become an MNC or a DC – but instead may have a continuous element to it that corresponds to firms

receiving various doses of internationalization (differing treatment amounts). In the absence of randomized trials with placebos and variable doses, we focus on observational data and use cross-sectional regressions to make inferences about the treatment effect (Angrist & Krueger, 2001). Continuing with our MNC vs DC example, a common approach is to estimate the relation between an observed firm characteristic (e.g., profits) and a measure of firm-level internationalization (either as a binary or continuous variable) across a broad sample of firms. Although this approach seems intuitively appealing, it creates an interpretation problem because it is difficult from this test to make causal inferences about the question of interest.

Using Observational Data: The Non-Random Sample

The challenge in using observational data and cross-sectional tests is that the individuals or firms in our treatment and control groups are not randomly selected. More specifically, in the cross-section of firms that we actually observe, firms emerge in distinct organizational and industry patterns. The variable of interest may even influence how firms emerge as multinational or domestic companies (the particular case of reverse causality). For instance, in comparing MNCs and DCs it seems plausible that more profitable firms can afford to develop international operations or that firm internationalization arises due to differences in managerial experience that also affect firm profitability. This creates a non-random treatment problem, and it is not one that simply inflates the “*t*-statistics.” Instead, we obtain inconsistent estimates of the impact of firm internationalization on firm profitability in our regressions, potentially leading to the rejection of true hypotheses or failure to reject false hypotheses (Woolridge, 2010). Thus, our empirical tests are distorted, and we may draw the wrong policy implications.

An illustration at the individual level often serves as the best example to highlight this non-random treatment problem. Consider an international business researcher who is interested in testing a program to help facilitate cross-cultural teamwork. For convenience, the researcher provides the training to a group of professors at the university where s/he is employed. One year later s/he observes faculty effectiveness in cross-cultural teams and compares this to cross-cultural team effectiveness in the general population. Specifically, s/he regresses the cross-cultural teamwork effectiveness

on faculty appointment and discovers that, consistent with a positive treatment, the university professors in the sample have greater cross-cultural team effectiveness than does the general population. The researcher then reports an effective cross-cultural teamwork effect with the treatment group and concludes that firms should consider approving the training program for workers in their companies.

Clearly, the above test procedure may give the wrong answer to our true question of interest because the university professors may have greater cultural awareness relative to the general population without receiving the treatment. The assignment to the treatment group was not random. Unfortunately, our typical regressions in international business are often even more problematic than this particular example. In this example, we have an idea about the direction of the bias because we have an educated guess about the nature of cultural sensitivity among university professors and the general population. Yet in most international business issues of interest, the direction of the bias is unknown. Moreover, in IB we are rarely able to give the treatments to the subjects even if they are randomly assigned; rather, the individuals or firms themselves often select to take the treatment (or not). In IB studies that do not take into account the non-random assignment problem, we routinely observe that *JIBS* reviewers recommend rejection.

The Prevalence of Non-Random Treatment Problems

Even a cursory glance at real world data indicates that firms do not emerge randomly or uniformly around the world. Similarly, individuals are not randomly assigned postings nor do they uniformly develop managerial expertise. As such, it is difficult to interpret the cross-sectional tests that we commonly employ in IB because our analysis violates the necessary conditions to make them a valid test (Roberts & Whited, 2011). Of course we are all familiar with this potential issue, which is often known under the broad title of “endogeneity.” A common misconception in the papers we review is that ruling out reverse causality solves the endogeneity problem. Unfortunately, the problem is more pervasive, and reverse causality is only one distinct case of the non-random treatment effect. Statisticians and econometricians have been discussing the issue for decades, and over the past several years their remedies have become quite common in empirical business research.



At the *Journal of International Business Studies (JIBS)* we find that the most successful studies in IB use the intuition and insights behind these methods in developing their research design to facilitate causal inference from their observational data.

In IB research the objective usually centers on providing evidence about the causal effects of some particular IB phenomenon. Because this research usually involves observational data, rather than random trials, the relevant goal in IB research design centers on the development of a test that best approximates a controlled experiment (Angrist & Krueger, 2001). As a result, studies that explicitly identify the source of variability in the dependent variable can develop appropriate tests that improve the researcher's ability to make causal inferences (Angrist & Pischke, 2008). This issue applies to a variety of approaches in IB research, not just the examples used for illustration in this essay. Research on the determinants of multinationality or studies that use data items such as individual patents also face this same endogeneity problem. Brenner (2011) exemplifies this approach to careful research design at *JIBS* in his study determining if resource advantages cause managers in firms with illegal international activity to cooperate with government prosecutions.

APPROXIMATING THE RANDOMIZED-CONTROLLED EXPERIMENT

This section provides a brief (and hopefully intuitive) explanation of some of the statistical remedies that IB scholars use in their cross-sectional tests. These short descriptions of several common methods for dealing with the non-random treatment problem are not the main focus of this essay (however, Roberts and Whited, 2011 provide a thorough analysis). Rather our emphasis is on the importance of careful research design that incorporates field research or institutional knowledge to develop tests with observational data to facilitate causal inference.

We primarily focus on non-random treatment (endogenous binary variable) rather than the continuous case endogeneity because it provides an intuitive framework for discussing strategies to exploit variation in the main independent variable to develop testable hypotheses. In this context, we believe that discussing some of the potential remedies may help researchers who submit articles to *JIBS* identify the manner in which their observational data can be used to approximate a randomized controlled experiment. In sum, we seek to

highlight the notion that IB research that recognizes the variability in the casual relationship and clearly identifies the strategy being used to approximate a controlled experiment has the best chance of success in the *JIBS* review process.

Control Variables and Fixed Effects

Theoretical predictions in international business research are often direct and straightforward, suggesting that internationalization causes some activity to occur. The simplest test in this circumstance is to focus on univariate statistical differences between the groups of interest (i.e., MNCs vs DCs). Yet we all appreciate that we must control for other individual or firm attributes to properly gauge the relation of interest. At the most basic level this occurs because we do not have randomized controlled experiments. In essence, the inclusion of control variables in a multivariate regression is an attempt to deal with the non-random nature of the treatment effect in our analysis. Unfortunately, in many circumstances, this control variable approach is insufficient to deal with the non-random treatment effect problems that we encounter. Potential sources of problems include the omission of some important variables, reverse causality, and measurement error in the variables of interest (Roberts & Whited, 2011). Thus, it appears to reviewers that this empirical approach is chosen because of its ease of use rather than because it emerged as a well-designed strategy to make the tests more like an experiment.

As the non-random treatment problem has been recognized for decades, several statistical techniques have been developed and included in standard statistical software packages to address these concerns (e.g., in STATA, SPSS). Perhaps one of the earliest empirical approaches to dealing with endogeneity is mechanical in nature, namely including unit-level fixed effects in the regression (Woolridge, 2010). Unit-level fixed effects, such as firm fixed effects, are well suited for circumstances with panel data and are strongly endorsed in many econometric textbooks (e.g., Greene, 2008). This approach essentially includes a dummy variable for each individual or firm and relies on changes of the causal variable within a given individual or firm.

Although fixed effects are easy to implement, their ability to effectively curb the non-random treatment problem depends on the nature of the endogeneity problem. Business researchers often

find that in a dynamic setting where firm characteristics slowly change over time, the use of fixed-effects removes the theoretical, cross-sectional variation of interest (Zhou, 2001).¹ The implication is that it can be difficult to find a meaningful relationship between the causal variable and the outcome variable with fixed effects, even if one truly exists. Evidence of causal relation with unit-level fixed effects can be quite compelling even though it may be difficult to interpret a lack of evidence.

Matching and Propensity Score Models

The matched sample approach essentially attempts to address the non-random treatment effect by creating a pseudo random sample. In many international business situations the most obvious approach to matching centers on firm size or industry in order to develop a sample where the treated and untreated firms overlap. Cassiman and Golovko (2011), for instance, use a matching model framework to address endogeneity in a *JIBS* study on innovation and exports. At the individual level, matching on education and experience represent common approaches. Intuitively, matching is a method to add control variables and allow the treatment effect to differ across firm type. Matching achieves this goal by eliminating firms from either the treatment or control group that do not have comparable firms in the other group and therefore minimize extrapolation (Angrist & Pischke, 2008). The cost of this improved estimation in terms of robustness is that such analysis is less generalizable to the broader universe.

In recent years, an approach labeled as “propensity score matching” has gained popularity because it allows a refined matching process along multiple individual or firm characteristics (Dehejia & Wahba, 1999). In the effort to create a matched sample in a study on MNCs for instance, researchers may attempt to effectively randomize the data by matching MNCs to DCs along several different dimensions such as total assets, industry, ownership structure, analyst following, and so forth. This particular approach of matching often uses a logit or probit model with the variable of interest (i.e., propensity to become an MNC) as the dependent variable. The researcher then matches MNCs to DCs based on their predicted propensity to become MNCs. Often these propensity score models use one-to-one firm matching and attempt to match firms on their predicted values (Caliendo & Kopeinig, 2008). Although one-to-one matching

exemplifies the spirit behind matching, alternative propensity score approaches such as one-to-many, kernel matching and reseeded may also be relevant.

Implementing a propensity score model with a binary treatment is straightforward. The first step is to predict the variable of interest for each individual or firm (i.e., predict their likelihood of becoming an MNC) based on multiple individual or firm characteristics. Second, using the predicted value for the variable of interest (i.e., chance of becoming an MNC) match individuals or firms with high and low values of the variable of interest (i.e., MNCs to DCs). Third, test the original equation of interest (i.e., profits in MNCs and DCs) using only the individuals or firms in the matched sample. In essence, this approach attempts to correct for the non-random treatment effect by matching a treated firm (or person) to an untreated firm which has similar characteristics.

Using our cross-cultural training example from earlier, each of our treated faculty members would be matched to someone in the general population, with similar age, gender, education, activity levels, marital status and so forth. Although this may not solve the non-treatment problem, it can potentially mitigate some of the associated problems. A limitation of the matching approach is that for a given propensity score, one might have a lot of the treated firms (e.g., the MNCs) but only a very few of the counter-factual firms to be matched (e.g., the DCs), making one-to-one matching difficult. A major strength of the matching approach is that it obliges us to explicitly identify the non-random component of the treatment effect and to determine the appropriate counter-factual firms (Heinrich, Maffioli, & Vazquez, 2010 provide a primer on using propensity score matching).

Instrumental Variable Approach

Another popular approach to dealing with endogeneity is to seek an exogenous proxy for the treatment or independent variable of interest (Larcker & Rusticus, 2010). This classic approach centers on finding a variable, called an instrument, which influences the independent variable (the right-hand-side variable) but appears unlikely to affect the dependent variable (the left-hand-side variable) except through its effect on the independent variable (Wintoki, Linck, & Netter, 2012). Cull, Haber, and Imai (2011) provide an example in *JIBS* of using an instrumental variable



approach in their analysis on related lending and the development of banking systems.

Focusing again on our cross-cultural example, an instrument would need to be something that is significantly related to the likelihood of being in the treatment group (i.e., related to being a university professor) but unlikely to be related to cultural sensitivity. For instance, a sudden and unexpected increase in the job market opportunities in the year a person received their graduate degree might be related to the decision to become a university professor but unrelated to the cultural sensitivity which often starts earlier in life. We then use this “instrument” to predict the treatment effect and use this predicted variable in the test of interest.

Ideally, an instrument should affect the main dependent variable through a single channel and in a single direction (Angrist & Krueger, 2001). Unfortunately, exogenous instruments are rare and difficult to find. However, as the instrumental variable is part of the standard toolkit of many business scholars, we often see attempts to use some other firm choice variable as an “instrument.” In a high percentage of the empirical papers, it appears that the chosen instrument(s) often turn out to be some other endogenous variable(s). It is common, for instance, to see leverage or firm size used as instruments, when these are obviously related to the dependent variable. This approach is usually justified by pointing to some other articles that also choose to use this particular endogenous instrument. Murray (2006) provides a detailed discussion of the problem with invalid instruments. Ultimately, the instrumental variable approach depends on the quality of the instrument being used.² Larcker and Rusticus (2010) provide a step-by-step guide to using instrumental variables.

Natural Experiments

Another approach to dealing with endogeneity centers on evaluating the variable of interest after some shock, such as the death of a CEO, a natural disaster, or a regulatory change. Using a specific intervention, such as a change in regulation, can be thought of as natural or quasi experiment (Bertrand, Duflo, & Mullainathan, 2004). The natural experiment approach uses the regulatory change as the treatment effect and allows the same firm or individual to be analyzed before and after the shock. To implement this approach one computes the difference between the variable of interest before the shock and after the shock in each firm affected by the shock or regulatory

change. Of course other things may be changing as well, so ideally we would like another set of firms or individuals that did not receive a shock to use as a control group. We then can compare the difference in the shock group to the difference in the non-shock group over the same time period. This difference-in-difference test provides a robust environment for evaluating cause and effect. The effectiveness of this approach depends on the exogeneity of the shock. For example, if a group of firms lobby to induce a regulatory change, then this regulatory change cannot really be considered an exogenous shock for these firms. In contrast, unexpected events like financial or political crisis can provide ideal test environments, especially when the shock and non-shock groups are similar along other firm or individual characteristics.

As an illustration, consider a researcher concerned about the impact of taxes on the investment strategies of multinational firms. A country changes its tax code in such way that taxes are increased for repatriated income, which primarily affects multinational firms. We then compare the investments by each MNC before and after the tax change. This difference provides an estimate of the effect of taxes on investments. Of course other issues in the economy may influence investments so we can compute this same difference in investments for domestic firms (who were unaffected by the change in the law). Computing the difference in these differences then provides a strong test of the effect of taxes on the investment decisions of multinational firms.

Regression Discontinuity Design

Another emerging technique for dealing with the non-random treatment effect centers on an approach labeled as regression discontinuity design (Lee & Lemieux, 2010). This method attempts to alleviate concerns about the non-random treatment effect by exploiting how people or firms become part of the treatment group. This approach focuses on identifying an observable characteristic that defines how someone or some firm becomes part of the treatment group and seeks to exploit the cut-off point. Essentially, the regression discontinuity method seeks to utilize the similarities of those individuals/firms just above and just below the cut-off point (Almond & Doyle, 2011). Thus, in a regression discontinuity design we would seek to compare firms (or persons) who were just above the cut-off point (became part of the treatment effect) to those who were just below the cut-off point.

As an example, assume we wish to compare the value of an expatriate posting, relative to a similar posting in a domestic subsidiary, on managerial career advancement in a sample of Finnish managers. Comparing the post placement salaries of the managers in both subsidiaries will provide a biased (upward) estimate of the foreign subsidiary posting because they are likely to be assigned to better managers. Even in the absence of the posting in the foreign subsidiary, the expat manager would, on average, likely earn a high wage in the future. To illustrate the process, assume that managers are assigned to subsidiary postings based on their IQ and that those with an IQ above 160 receive the expatriate posting and those between 140 and 160 receive assignments in the domestic subsidiary (with mean IQ of 100 and standard deviation of 15). Then, even though the managers are not randomly assigned to subsidiary postings, we may be able to extract the expatriate treatment effect because IQ data is available for male Finnish citizens as part of their compulsory military service.³ Presumably the ability of those with an IQ of 159 does not differ that much from those with an IQ of 161.

To evaluate the value of an expat posting we might regress post assignment pay on an indicator variable for expat posting by using the subset of the managers with IQs between 158 and 163. The counter-factual or control group is comprised of the managers with 158–160 IQs, while the treatment group is comprised of the expat managers with IQs between 161 and 163. The difference in pay between these two similar groups will be captured by the coefficient estimate on expatriate posting. In a sense, this approach endeavors to randomize the treatment group in a similar spirit to the propensity score model by suggesting the appropriate control and treatment groups are those on either side of the cut-off point. As such, this approach also represents a subset analysis.

A GUIDEPOST TO RESEARCH DESIGN: THE ROLE OF THEORY

The empirical approaches that seek to analyze data using standard regressions with matched samples, instrumental variables, natural experiments or regression discontinuity designs are valid and relevant. Unfortunately, we frequently use these statistical techniques as crutches or substitutes for critically thinking about the problem of interest, resulting in dubious analyses (Thomas, Cuervo-Cazurra, & Brannen, 2011). In our haste to discover

the truth, we often seek to let the data speak by running regressions and then fashioning a story around the results. Yet this approach intensifies and exacerbates the non-random treatment problem, leading to deceptive results and improper policy implications. The systematic manner in which the underlying data emerge needs to be considered before the first test is performed (Heckman & Urzua, 2009). Thus, the first step in developing our hypotheses is to identify how firms/individuals are assigned to the treatment group and why this assignment occurs (Roberts & Whited, 2011).

While this first step sounds simple, we often find this critical step is skipped in the papers submitted to *JIBS*. Take an example of a study of the relationship between a country's legal system and the behavior of firms in a country (e.g., propensity of foreign direct investment or FDI) based on the panel data of multiple countries and firms. A country's legal system is determined by its resources, history, culture, industry structure and so on, and it does not change quickly. Therefore the legal system affects firm behavior, not the other way around, and hypotheses should be developed in that direction. In the long run, however, the firms in a country (or even foreign firms) can affect the country's legal system. For example, more profitable firms might demand stricter intellectual property rights protection and their taxes may help fund the legal system, so one might be interested in investigating this kind of phenomenon. If one wants to study how firm behavior affects legal systems, then the potential for non-random legal institution assignment should be fully examined and incorporated in hypothesis development. In this case, one cannot delegate the examination of potential reverse causality (i.e., legal system affects firm behavior) to statistical tests.

A Systematic Approach

One formal approach to dealing with the non-random treatment effect centers on developing a structural model. Structural models provide rigid and explicit equations of individual or firm behavior that rely on idealistic assumptions.⁴ Although structural models are often couched in technical jargon, the intuition behind using them suggests a simple framework for developing the theoretical underpinnings of the eventual empirical specification. Fundamentally, one should think about how the observed variations in the right-hand-side variable of interest may have



emerged. The scholar's institutional knowledge and ideas about how the treatment decision emerges are critical components to sound empirical research design (Angrist & Pischke, 2008). As we are unable to randomly assign firms into the treatment/control groups, understanding how the firms were initially assigned to the treatment or control group is essential to developing testable hypotheses.

Consider an example regarding the determinants of FDI. An IB scholar might be interested in evaluating the idea that firm-level FDI is driven by firms seeking to find low cost employees. Accounts in the business press routinely describe investment in China and job migration of the US in this fashion. One approach to test this idea would be to compare FDI within a country, across different states/provinces, based on the average wage rate. Alternatively, one could make the same sort of comparison across multiple countries or geographic regions. A typical premise to test this maintained hypothesis might be: FDI is negatively related to wages. Basile (2004) provides such a test in the context of FDI across Italy using foreign acquisitions. Specifically, evidence is found to suggest foreign direct investment is positively related to wage rates. One might be tempted to conclude that our theoretical prediction was incorrect; instead we found that firms were targeting FDI in provinces with high wages.

Yet a finding of a positive relation between wages and FDI may stem from the non-random sample that we used. Recall that our tests are based on the premise that we randomly assigned high wages to some countries/provinces and low wages to others. Ideally, the wage rate is supposed to be randomly assigned across countries to generate a reliable test. Of course, wage rates are not randomly distributed across countries/provinces but instead may arise due to differences in human and physical capital. Thus, our hypothesis and research design needs to incorporate the notion that wage rates are not exogenous. For instance, it may be that wage rates are a function of education and experience, suggesting that low wage environments may have limited human capital. In terms of identifying the relevant control group this non-random component needs to be incorporated into the hypothesis development.

Focusing on the theoretical development in this FDI story, our hypotheses need to explicitly acknowledge that FDI should be negatively related to wages for a given level of human capital. This type of hypothesis might naturally lead to the

construction of a propensity score matched sample of workers from high and low wage provinces with similar levels of human capital in order to identify the wage rate differential. Ultimately, none of the procedures developed by econometricians are magic pills (Roberts & Whited, 2011). Instead, they all highlight the need for careful theoretical development that leads to the proper identification of the relevant control group as the best alternative to randomized controlled experiments.

Research Implications

The success of empirical business research over the past three decades is based on simple, straightforward theories that provide qualitative predictions and conform to observed real world phenomena. Yet we still need to incorporate the notion that the independent variable is unlikely to be randomly distributed across firms. Generally speaking, the theoretical predictions that we develop should incorporate, by design, the non-random component of our right-hand-side variable. Reinterpreting Marschak (1953), it is not necessary or desirable to fully specify a structural model of the dependent variable, but one does need to consider the fundamental economic issues that lead to non-random assignments of the treatment effect among the firms in the sample. In the absence of a randomized controlled experiment, we need to incorporate the non-random assignment into the treatment group in our research design to improve causal inference.

More fundamentally, the strength of IB research depends on the ability to identify the main theoretical mechanisms by which the dependent variable arises. These mechanisms can be identified using both qualitative and quantitative methods. One approach that seems to be gaining ground in more recent years centers on combining field research (anecdotal or systematic qualitative) and quantitative analysis. Because this approach relies on insights from the insiders – managers and employees of the firm – to inform econometric analysis, it is sometimes known as “insider research” (Ichniowski & Shaw, 2009). In insider research, the rich micro data collected through field interviews help identify the behavioral mechanisms that explain how the treatment (for instance, a certain type of management work practice) affects firm performance including productivity and profitability (Siegel & Larson, 2009). Similarly, speaking with professionals about the nature of causality for researchers using secondary data can also be useful.

By helping the researcher to model the adoption of treatments more accurately, insider research helps identify any selection bias in the estimation of the treatment effect. The key issues though center on developing a strong theoretical argument for how a phenomenon causes a particular effect and how that phenomenon emerged among the observations in the sample.

Across multiple international business subfields, we find that researchers who carefully consider how a phenomenon arose in the cross section are often the most successful at *JIBS*. Following these examples we encourage researchers to speak to managers, market participants, bankers or consultants in the area to obtain the institutional details that are crucial to understanding the nature of causality in a particular phenomenon. The acid test is whether the research design in an empirical study with observational data is the one that best approximates a randomized-controlled experiment for the hypothesis of interest.

REFERENCES

- Almond, D., & Doyle, J. J. 2011. After midnight: A regression discontinuity design in length of postpartum hospital stays. *American Economic Journal: Economic Policy*, 3(3): 1–34.
- Angrist, J., & Krueger, A. 2001. Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, 15(4): 69–85.
- Angrist, J., & Pischke, J. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.
- Basile, R. 2004. Acquisitions versus greenfield investment: The location of foreign manufacturers in Italy. *Regional Science and Urban Economics*, 34(1): 3–25.
- Bertrand, M., Duflo, E., & Mullainathan, S. 2004. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1): 249–275.
- Brenner, S. 2011. Self-disclosure at international cartels. *Journal of International Business Studies*, 42(2): 221–234.
- Caliendo, M., & Kopeinig, S. 2008. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1): 31–72.
- Cassiman, B., & Golovko, E. 2011. Innovation and internationalization through exports. *Journal of International Business Studies*, 42(1): 56–75.
- Cull, R., Haber, S., & Imai, M. 2011. Related lending and banking development. *Journal of International Business Studies*, 42(3): 406–426.
- Dehejia, R., & Wahba, S. 1999. Causal effects in non-experimental studies: Re-evaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448): 1053–1062.
- Greene, W. 2008. *Econometric analysis* (6th edn) Upper Saddle River, NJ: Prentice Hall.
- Heckman, J., & Urzua, S. 2009. *Comparing IV and structural models: What simple IV can and cannot identify*, Working Paper, National Bureau of Economic Research.
- Heinrich, C., Maffioli, A., & Vazquez, G. 2010. *A primer for applying propensity-score matching*, Technical Note IDB-TN-161, Inter-American Development Bank.
- Hoekstra, M. 2009. The effect of attending the flagship state university on earnings: A discontinuity based approach. *Review of Economics and Statistics*, 91(4): 717–724.
- Ichniowski, C., & Shaw, K. 2009. *Insider econometrics: Empirical studies of how management matters*, Working Paper 15618, National Bureau of Economic Research.
- Larcker, D., & Rusticus, T. 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics*, 49(3): 186–205.
- Lee, D., & Lemieux, T. 2010. Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2): 281–355.
- Marschak, J. 1953. Economic measurements for policy and prediction. In W. Hood & T. Koopmans (Eds), *Studies in econometric method*: 1–26. New York: Wiley.
- Murray, M. 2006. Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives*, 20(4): 111–132.
- Roberts, M., & Whited, T. 2011. *Endogeneity in corporate finance*, Working Paper, Wharton.
- Siegel, J., & Larson, B. Z. 2009. Labor market institutions and global strategic adaptation: Evidence from Lincoln Electric. *Management Science*, 55(9): 1527–1546.
- Thomas, D. C., Cuervo-Cazurra, A., & Brannen, M. Y. 2011. From the editors: Explaining theoretical relationships in international business research: Focusing on the arrows, NOT the boxes. *Journal of International Business Studies*, 42(9): 1073–1078.
- Welch, I. 2010. *A critique of quantitative structural models in corporate finance*, Working Paper, Brown University.
- Wintoki, M., Linck, J., & Netter, J. 2012. Endogeneity and the dynamics of internal corporate governance. *Journal of Financial Economics*, forthcoming.
- Wooldridge, J. 2010. *Econometric analysis of cross section and panel data*, (2nd edn) Cambridge, MA: MIT Press.
- Zhou, X. 2001. Understanding the determinants of managerial ownership and the link between ownership and performance: Comment. *Journal of Financial Economics*, 62(3): 559–571.

NOTES

¹Specifically, the fixed effect absorbs the time-invariant characteristics of the firm, which mitigates endogeneity but also reduces our ability to study the effects of some time-invariant variables of interest.

²Using our cross-cultural training example again, a poor choice for predicting treatment might be religious tolerance. Although religious tolerance might be related to university employment, it may also be related to cross-cultural sensitivity.

³The subsidiary assignment could be done with some other set of observable criteria such as employee rankings or assessments. IQ, however, provides a useful illustration due to its fine gradation and familiarity to academics. Hoekstra (2009) provides an example of using regression discontinuity design to evaluate flagship university attendance on salaries.

⁴Consequently, most business questions do not lend themselves to using formal quantitative structural models that focus on complex predictions (see Welch, 2010).