

Endogeneity in Empirical Corporate Finance*

Michael R. Roberts

The Wharton School, University of Pennsylvania and NBER

Toni M. Whited

Simon Graduate School of Business, University of Rochester

First Draft: January 24, 2011

Current Draft: October 5, 2012

*Roberts is from the Finance Department, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6367. Email: mrobert@wharton.upenn.edu. Whited is from the Simon School of Business, University of Rochester, Rochester, NY 14627. Email: toni.whited@simon.rochester.edu. We thank the editors, George Constantinides, Milt Harris, and Rene Stulz for comments and suggestions. We also thank Don Bowen, Murray Frank, Todd Gormley, Mancy Luo, Andrew Mackinlay, Phillip Schnabl, Ken Singleton, Roberto Wessels, Shan Zhao, Heqing Zhu, the students of Finance 926 at the Wharton School, and the students of Finance 534 at the Simon School for helpful comments and suggestions.

Abstract

This chapter discusses how applied researchers in corporate finance can address endogeneity concerns. We begin by reviewing the sources of endogeneity—omitted variables, simultaneity, and measurement error—and their implications for inference. We then discuss in detail a number of econometric techniques aimed at addressing endogeneity problems including: instrumental variables, difference-in-differences estimators, regression discontinuity design, matching methods, panel data methods, and higher order moments estimators. The unifying themes of our discussion are the emphasis on intuition and the applications to corporate finance.

Keywords: Instrumental Variables, Difference-in-Differences Estimators, Regression Discontinuity Designs, Matching Estimators, Measurement Error

J.E.L. Codes: G3, C21, C23, C26

Contents

1. Introduction	6
2. The Causes and Consequences of Endogeneity	7
2.1 Regression Framework	8
2.1.1 Omitted Variables	10
2.1.2 Simultaneity	11
2.1.3 Measurement Error	13
2.2 Potential Outcomes and Treatment Effects	17
2.2.1 Notation and Framework	17
2.2.2 An Example	21
2.2.3 The Link to Regression and Endogeneity	22
2.2.4 Heterogeneous Effects	23
2.3 Identifying and Discussing the Endogeneity Problem	24
3. Instrumental Variables	24
3.1 What are Valid Instruments?	24
3.2 Estimation	26
3.3 Where do Valid Instruments Come From? Some Examples	27
3.4 So Called Tests of Instrument Validity	28
3.5 The Problem of Weak Instruments	29
3.6 Lagged Instruments	31
3.7 Limitations of Instrumental Variables	32
4. Difference-in-Differences Estimators	34
4.1 Single Cross-Sectional Differences After Treatment	34
4.2 Single Time-Series Difference Before and After Treatment	35

4.3	Double Difference Estimator: Difference-in-Differences (DD)	38
4.3.1	Revisiting the Single Difference Estimators	40
4.3.2	Model Extensions	40
4.3.3	The Key Identifying Assumption for DD	41
4.4	Checking Internal Validity	44
4.5	Further Reading	47
5.	Regression Discontinuity Design	47
5.1	Sharp RDD	48
5.1.1	Identifying Treatment Effects	50
5.2	Fuzzy RDD	52
5.2.1	Identifying Treatment Effects	54
5.3	Graphical Analysis	56
5.4	Estimation	58
5.4.1	Sharp RDD	58
5.4.2	Fuzzy RDD	61
5.4.3	Semiparametric Alternatives	62
5.5	Checking Internal Validity	63
5.5.1	Manipulation	63
5.5.2	Balancing Tests and Covariates	65
5.5.3	Falsification Tests	66
6.	Matching Methods	66
6.1	Treatment Effects and Identification Assumptions	67
6.2	The Propensity Score	68
6.3	Matching on Covariates and the Propensity Score	69
6.4	Practical Considerations	71

6.4.1	Assessing Unconfoundedness and Overlap	71
6.4.2	Choice of Distance Metric	73
6.4.3	How to Estimate the Propensity Score?	73
6.4.4	How Many Matches?	74
6.4.5	Match with or without Replacement?	74
6.4.6	Which Covariates?	75
6.4.7	Matches for Whom?	75
7.	Panel Data Methods	76
7.1	Fixed and Random Effects	76
8.	Econometric Solutions to Measurement Error	78
8.1	Instrumental Variables	79
8.2	High Order Moment Estimators	81
8.3	Reverse Regression Bounds	83
8.4	Avoiding Proxies and Using Proxies Wisely	85
9.	Conclusion	85

1. Introduction

Arguably, the most important and pervasive issue confronting studies in empirical corporate finance is endogeneity, which we can loosely define as a correlation between the explanatory variables and the error term in a regression. Endogeneity leads to biased and inconsistent parameter estimates that make reliable inference virtually impossible. In many cases, endogeneity can be severe enough to reverse even qualitative inference. Yet, the combination of complex decision processes facing firms and limited information available to researchers ensures that endogeneity concerns are present in every study. These facts raise the question: how can corporate finance researchers address endogeneity concerns? Our goal is to answer this question.

However, as stated, our goal is overly ambitious for a single survey paper. As such, we focus our attention on providing a practical guide and starting point for addressing endogeneity issues encountered in corporate finance. Recognition of endogeneity issues has increased noticeably over the last decade, along with the use of econometric techniques targeting these issues. Although this trend is encouraging, there have been some growing pains as the field learns new econometric techniques and translates them to corporate finance settings. As such, we note potential pitfalls when discussing techniques and their application. Further, we emphasize the importance of designing studies with a tight connection between the economic question under study and the econometrics used to answer the question.

We begin by briefly reviewing the sources of endogeneity—omitted variables, simultaneity, and measurement error—and their implications for inference. While standard fare in most econometrics textbooks, our discussion of these issues focuses on their manifestation in corporate finance settings. This discussion lays the groundwork for understanding how to address endogeneity problems.

We then review a number of econometric techniques aimed at addressing endogeneity problems. These techniques can be broadly classified into two categories. The first category includes techniques that rely on a clear source of exogenous variation for identifying the coefficients of interest. Examples of these techniques include instrumental variables, difference-in-differences estimators, and regression discontinuity design. The second category includes techniques that rely more heavily on modeling assumptions, as opposed to a clear source of exogenous variation. Examples of these techniques include panel data methods (e.g., fixed and random effects), matching methods, and measurement error methods.

In discussing these techniques, we emphasize intuition and proper application in the context of corporate finance. For technical details and formal proofs of many results, we refer

readers to the appropriate econometric references. In doing so, we hope to provide empirical researchers in corporate finance not only with a set of tools, but also an instruction manual for the proper use of these tools.

Space constraints necessitate several compromises. Our discussion of selection problems is confined to that associated with non-random assignment and the estimation of causal effects. A broader treatment of sample selection issues is contained in Li and Prabhala (2007).¹ We also do not discuss structural estimation, which relies on an explicit theoretical model to impose identifying restrictions. Most of our attention is on linear models and nonparametric estimators that have begun to appear in corporate finance applications. Finally, we avoid details associated with standard error computations and, instead, refer the reader to the relevant econometrics literature and the recent study by Petersen (2009).

The remainder of the paper proceeds as follows. Section 2 begins by presenting the basic empirical framework and notation used in this paper. We discuss the causes and consequences of endogeneity using a variety of examples from corporate finance. Additionally, we introduce the potential outcomes notation used throughout the econometric literature examining treatment effects and discuss its link to linear regressions. In doing so, we hope to provide an introduction to the econometrics literature that will aid and encourage readers to stay abreast of econometric developments.

Sections 3 through 5 discuss techniques falling in the first category mentioned above: instrumental variables, difference-in-differences estimators, and regression discontinuity designs. Sections 6 through 8 discuss techniques from the second category: matching methods, panel data methods, and measurement error methods. Section 9 concludes with our thoughts on the subjectivity inherent in addressing endogeneity and several practical considerations. We have done our best to make each section self-contained in order to make the chapter readable in a nonlinear or piecemeal fashion.

2. The Causes and Consequences of Endogeneity

The first step in addressing endogeneity is identifying the problem.² More precisely, researchers must make clear which variable(s) are endogenous and why they are endogenous.

¹A more narrow treatment of econometric issues in corporate governance can be found in Bhagat and Jeffries (2005).

²As Wooldridge notes, endogenous variables traditionally refer to those variables determined within the context of a model. Our definition of correlation between an explanatory variable and the error term in a regression is broader.

Only after doing so can one hope to devise an empirical strategy that appropriately addresses this problem. The goal of this section is to aid in this initial step.

The first part of this section focuses on endogeneity in the context of a single equation linear regression—the workhorse of the empirical corporate finance literature. The second part introduces treatment effects and potential outcomes notation. This literature that studies the identification of causal effects is now pervasive in several fields of economics (e.g., econometrics, labor, development, public finance). Understanding potential outcomes and treatment effects is now a prerequisite for a thorough understanding of several modern econometric techniques, such as regression discontinuity design and matching. More importantly, an understanding of this framework is useful for empirical corporate finance studies that seek to identify the causal effects of binary variables on corporate behavior.

We follow closely the notation and conventions in Wooldridge (2002), to which we refer the reader for further detail.

2.1 Regression Framework

In population form, the single equation linear model is

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \tag{1}$$

where y is a scalar random variable referred to as the outcome or dependent variable, (x_1, \dots, x_k) are scalar random variables referred to as explanatory variables or covariates, u is the unobservable random error or disturbance term, and $(\beta_0, \dots, \beta_k)$ are constant parameters to be estimated.

The key assumptions needed for OLS to produce consistent estimates of the parameters are the following:

1. a random sample of observations on y and (x_1, \dots, x_k) ,
2. a mean zero error term (i.e., $E(u) = 0$),
3. no linear relationships among the explanatory variables (i.e., no perfect collinearity so that $\text{rank}(X'X) = k$, where $X = (1, x_1, \dots, x_k)$ is a $1 \times (k + 1)$ vector), and
4. an error term that is uncorrelated with each explanatory variable (i.e., $\text{cov}(x_j, u) = 0$ for $j = 1, \dots, k$).

For unbiased estimates, one must replace assumption 4. with:

- 4a. an error term with zero mean conditional on the explanatory variables (i.e., $E(u|X) = 0$).

Assumption 4a is weaker than statistical independence between the regressors and error term, but stronger than zero correlation. Conditions 1 through 4 also ensure that OLS identifies the parameter vector, which in this linear setting implies that the parameters can be written in terms of population moments of (y, X) .³

A couple of comments concerning these assumptions are in order. The first assumption can be weakened. One need assume only that the error term is independent of the sample selection mechanism conditional on the covariates. The second assumption is automatically satisfied by the inclusion of an intercept among the regressors.⁴ Strict violation of the third assumption can be detected when the design matrix is not invertible. Practically speaking, most computer programs will recognize and address this problem by imposing the necessary coefficient restrictions to ensure a full rank design matrix, X . However, one should not rely on the computer to detect this failure since the restrictions, which have implications for interpretation of the coefficients, can be arbitrary.

Assumption 4 (or 4a) should be the focus of most research designs because violation of this assumption is the primary cause of inference problems. Yet, this condition is empirically untestable because one cannot observe u . We repeat *there is no way to empirically test whether a variable is correlated with the regression error term because the error term is unobservable*. Consequently, there is no way to statistically ensure that an endogeneity problem has been solved.

In the following subsections, each of the three causes of endogeneity maintains Assumptions 1 through 3. We introduce specification changes to Eqn (1) that alter the error term in a manner that violates Assumption 4 and, therefore, introduces an endogeneity problem.

³To see this statistical identification, write Eqn (1) as $y = XB + u$, where $B = (\beta_0, \beta_1, \dots, \beta_k)'$ and $X = (1, x_1, \dots, x_k)$. Premultiply this equation by X' and take expectations so that $E(X'y) = E(X'X)B$. Solving for B yields $B = E(X'X)^{-1}E(X'y)$. In order for this equation to have a unique solution, assumptions 3 and 4 (or 4a) must hold.

⁴Assume that $E(u) = r \neq 0$. We can rewrite $u = r + w$, where $E(w) = 0$. The regression is then $y = \alpha + \beta_1x_1 + \dots + \beta_kx_k + w$, where $\alpha = (\beta_0 + r)$. Thus, a nonzero mean for the error term simply gets absorbed by the intercept.

2.1.1 Omitted Variables

Omitted variables refer to those variables that should be included in the vector of explanatory variables, but for various reasons are not. This problem is particularly severe in corporate finance. The objects of study (firms or CEOs, for example) are heterogeneous along many different dimensions, most of which are difficult to observe. For example, executive compensation depends on executives' abilities, which are difficult to quantify much less observe. Likewise, financing frictions such as information asymmetry and incentive conflicts among a firms' stakeholders are both theoretically important determinants of corporate financial and investment policies; yet, both frictions are difficult to quantify and observe. More broadly, most corporate decisions are based on both public and nonpublic information, suggesting that a number of factors relevant for corporate behavior are unobservable to econometricians.

The inability to observe these determinants means that instead of appearing among the explanatory variables, X , these omitted variables appear in the error term, u . If these omitted variables are uncorrelated with the included explanatory variables, then there is no problem for inference; the estimated coefficients are consistent and, under the stronger assumption of zero conditional mean, unbiased. If the two sets of variables are correlated, then there is an endogeneity problem that causes inference to break down.

To see precisely how inference breaks down, assume that the true economic relation is given by

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \gamma w + u, \quad (2)$$

where w is an unobservable explanatory variable and γ its coefficient. The estimable population regression is

$$y = \beta_0 + \beta x_1 + \cdots + \beta_k x_k + v, \quad (3)$$

where $v = \gamma w + u$ is the composite error term. We can assume without loss of generality that w has zero mean since any nonzero mean will simply be subsumed by the intercept.

If the omitted variable w is correlated with any of the explanatory variables, (x_1, \dots, x_k) , then the composite error term v is correlated with the explanatory variables. In this case, OLS estimation of Eqn (3) will typically produce inconsistent estimates of *all* of the elements of β . When only one variable, say x_j , is correlated with the omitted variable, it is possible to understand the direction and magnitude of the asymptotic bias. However, this situation is highly unlikely, especially in corporate finance applications. Thus, most researchers implicitly assume that all of the other explanatory variables are partially uncorrelated with the omitted variable. In other words, a regression of the omitted variable on all of the explanatory variables would produce zero coefficients for each variable except x_j . In this case, the

probability limit for the estimate of β_l (denoted $\hat{\beta}_l$) is equal to β_l for $l \neq j$, and for β_j

$$\text{plim } \hat{\beta}_j = \beta_j + \gamma\phi_j, j = 1, \dots, k \quad (4)$$

where $\phi_j = \text{cov}(x_j, w)/\text{Var}(x_j)$.

Eqn (4) is useful for understanding the direction and potential magnitude of any omitted variables inconsistency. This equation shows that the OLS estimate of the endogenous variable's coefficient converges to the true value, β_j , plus a bias term as the sample size increases. The bias term is equal to the product of the effect of the omitted variable on the outcome variable, γ , and the effect of the omitted variable on the included variable, ϕ_j . If w and x_j are uncorrelated, then $\phi_j = 0$ and OLS is consistent. If w and x_j are correlated, then OLS is inconsistent. If γ and ϕ_j have the same sign—positive or negative—then the asymptotic bias is positive. With different signs, the asymptotic bias is negative.

Eqn (4) in conjunction with economic theory can be used to gauge the importance and direction of omitted variables biases. For example, firm size is a common determinant in CEO compensation studies (e.g., Core, Guay, and Larcker, 2008). If larger firms are more difficult to manage, and therefore require more skilled managers (Gabaix and Landier, 2008), then firm size is endogenous because managerial ability, which is unobservable, is in the error term and is correlated with an included regressor, firm size. Using the notation above, y is a measure of executive compensation, x is a measure of firm size, and w is a measure of executive ability. The bias in the estimated firm size coefficient will likely be positive, assuming that the partial correlation between ability and compensation (γ) is positive, and that the partial correlation between ability and firm size (ϕ_j) is also positive. (By partial correlation, we mean the appropriate regression coefficient.)

2.1.2 Simultaneity

Simultaneity bias occurs when y and one or more of the x 's are determined in equilibrium so that it can plausibly be argued either that x_k causes y or that y causes x_k . For example, in a regression of a value multiple (such as market-to-book) on an index of antitakeover provisions, the usual result is a negative coefficient on the index. However, this result does not imply that the presence of antitakeover provisions leads to a loss in firm value. It is also possible that managers of low-value firms adopt antitakeover provisions in order to entrench themselves.⁵

⁵See Schoar and Washington (2010) for a recent discussion of the endogenous nature of governance structures with respect to firm value.

Most prominently, simultaneity bias also arises when estimating demand or supply curves. For example, suppose y in Eqn (1) is the interest rate charged on a loan, and suppose that x is the quantity of the loan demanded. In equilibrium, this quantity is also the quantity supplied, which implies that in any data set of loan rates and loan quantities, some of these data points are predominantly the product of demand shifts, and others are predominantly the product of supply shifts. The coefficient estimate on x could be either positive or negative, depending on the relative elasticities of the supply and demand curves as well as the relative variation in the two curves.⁶

To illustrate simultaneity bias, we simplify the example of the effects of antitakeover provisions on firm value, and we consider a case in which Eqn (1) contains only one explanatory variable, x , in which both y and x have zero means, in which y and x are determined jointly as follows:

$$y = \beta x + u, \tag{5}$$

$$x = \alpha y + v, \tag{6}$$

and with u uncorrelated with v . We can think of y as the market-to-book ratio and x as a measure of antitakeover provisions. To derive the bias from estimating Eqn (5) by OLS, we can write the population estimate of the slope coefficient of Eqn (5) as

$$\begin{aligned} \hat{\beta} &= \frac{\text{cov}(x, y)}{\text{var}(x)} \\ &= \frac{\text{cov}(x, \beta x + u)}{\text{var}(x)} \\ &= \beta + \frac{\text{cov}(x, u)}{\text{var}(x)} \end{aligned}$$

Using Eqns (5) and (6) to solve for x in terms of u and v , we can write the last bias term as

$$\frac{\text{cov}(x, u)}{\text{var}(x)} = \frac{\alpha(1 - \alpha\beta)\text{var}(u)}{\alpha^2\text{var}(u) + \text{var}(v)}$$

This example illustrates the general principle that, unlike omitted variables bias, simultaneity bias is difficult to sign because it depends on the relative magnitudes of different effects, which cannot be known a priori.

⁶See Ivashina (2009) for a related examination of the role of lead bank loan shares on interest rate spreads. Likewise, Murfin (2010) attempts to identify supply-side determinants of loan contract features—covenant strictness—using an instrumental variables approach.

2.1.3 Measurement Error

Most empirical studies in corporate finance use proxies for unobservable or difficult to quantify variables. Any discrepancy between the true variable of interest and the proxy leads to measurement error. These discrepancies arise not only because data collectors record variables incorrectly but also because of conceptual differences between proxies and their unobservable counterparts. When variables are measured imperfectly, the measurement error becomes part of the regression error. The impact of this error on coefficient estimates, not surprisingly, depends crucially on its statistical properties. As the following discussion will make clear, measurement error does *not* always result in an attenuation bias in the estimated coefficient—the default assumption in many empirical corporate finance studies. Rather, the implications are more subtle.

Measurement Error in the Dependent Variable

Consider the situation in which the dependent variable is measured with error. Capital structure theories such as Fischer, Heinkel, and Zechner (1989) and Leland (1994) consider a main variable of interest to be the market leverage ratio, which is the ratio of the market value of debt to the market value of the firm (debt plus equity). While the market value of equity is fairly easy to measure, the market value of debt is more difficult. Most debt is privately held by banks and other financial institutions, so there is no observable market value. Most public debt is infrequently traded, leading to stale quotes as proxies for market values. As such, empirical studies often use book debt values in their place, a situation that creates a wedge between the empirical measure and the true economic measure. For the same reason, measures of firm, as opposed to shareholder, value face measurement difficulties. Total compensation for executives can also be difficult to measure. Stock options often vest over time and are valued using an approximation, such as Black-Scholes (Core, Guay, and Larcker, 2008).

What are the implications of measurement error in the dependent variable? Consider the population model

$$y^* = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u,$$

where y^* is an unobservable measure and y is the observable version of or proxy for y^* . The difference between the two is defined as $w \equiv y - y^*$. The estimable model is

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + v, \tag{7}$$

where $v = w + u$ is the composite error term. Without loss of generality, we can assume that

w , like u , has a zero mean so that v has a zero mean.⁷

The similarity between Eqns (7) and (3) is intentional. The statistical implications of measurement error in the dependent variable are similar to those of an omitted variable. If the measurement error is uncorrelated with the explanatory variables, then OLS estimation of Eqn (7) produces consistent estimates; if correlated, then OLS estimates are inconsistent. Most studies assume the former, in which case the only impact of measurement error in the dependent variable on the regression is on the error variance and parameter covariance matrix.⁸

Returning to the corporate leverage example above, what are the implications of measurement error in the value of firm debt? As firms become more distressed, the market value of debt will tend to fall by more than the book value. Yet, several determinants of capital structure, such as profitability, are correlated with distress. Ignoring any correlation between the measurement error and other explanatory variables allows us to use Eqn (4) to show that this form of measurement error would impart a downward bias on the OLS estimate of the profitability coefficient.⁹

Measurement Error in the Independent Variable

Next, consider measurement error in the explanatory variables. Perhaps the most recognized example is found in the investment literature. Theoretically, marginal q is a sufficient statistic for investment (Hayashi, 1982). Empirically, marginal q is difficult to measure, and so a number of proxies have been used, most of which are an attempt to measure Tobin's q —the market value of assets divided by their replacement value. Likewise, the capital structure literature is littered with proxies for everything from the probability of default, to the tax benefits of debt, to the liquidation value of assets. Studies of corporate governance also rely greatly on proxies. Governance is itself a nebulous concept with a variety of different facets. Variables such as an antitakeover provision index or the presence of a large blockholder are unlikely sufficient statistics for corporate governance, which includes the strength of board oversight among other things.

What are the implications of measurement error in an independent variable? Assume

⁷In general, biased measurement in the form of a nonzero mean for w only has consequences for the intercept of the regression, just like a nonzero mean error term.

⁸If u and w are uncorrelated, then measurement error in the dependent variable increases the error variance since $\sigma_v^2 = \sigma_w^2 + \sigma_u^2 > \sigma_u^2$. If they are correlated, then the impact depends on the sign and magnitude of the covariance term.

⁹The partial correlation between the measurement error in leverage and book leverage (γ) is positive: measurement error is larger at higher levels of leverage. The partial correlation between the measurement error in leverage and profitability (ϕ_j) is negative: measurement error is larger at lower levels of profits.

the population model is

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k^* + u, \quad (8)$$

where x_k^* is an unobservable measure and x_k is its observable proxy. We assume that u is uncorrelated with all of the explanatory variables in Eqn (8), $(x_1, \dots, x_{k-1}, x_k^*)$, as well as the observable proxy x_k . Define the measurement error to be $w \equiv x_k - x_k^*$, which is assumed to have zero mean without loss of generality. The estimable model is

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + v, \quad (9)$$

where $v = u - \beta_k w$ is the composite error term.

Again, the similarity between Eqns (9) and (3) is intentional. As long as w is uncorrelated with each x_j , OLS will produce consistent estimates since a maintained assumption is that u is uncorrelated with all of the explanatory variables — observed and unobserved. In particular, if the measurement error w is uncorrelated with the observed measure x_k , then none of the conditions for the consistency of OLS are violated. What is affected is the variance of the error term, which changes from $\text{var}(u) = \sigma_u^2$ to $\text{var}(u - \beta_k w) = \sigma_u^2 + \beta_k^2 \sigma_w^2 - 2\beta_k \sigma_{uw}$. If u and w are uncorrelated, then the regression error variance increases along with the estimated standard errors, all else equal.

The more common assumption, referred to as the classical errors-in-variables assumption, is that the measurement error is uncorrelated with the unobserved explanatory variable, x_k^* . This assumption implies that w must be correlated with x_k since $\text{cov}(x_k, w) = E(x_k w) = E(x_k^* w) + E(w^2) = \sigma_w^2$. Thus, x_k and the composite error v from Eqn (9) are correlated, violating the orthogonality condition (assumption 4). This particular error-covariate correlation means that OLS produces the familiar attenuation bias on the coefficient of the mismeasured regressor.

The probability limit of the coefficient on the tainted variable can be characterized as:

$$\text{plim} \hat{\beta}_k = \beta_k \left(\frac{\sigma_r^2}{\sigma_r^2 + \sigma_w^2} \right) \quad (10)$$

where σ_r^2 is the error variance from a linear regression of x_k^* on (x_1, \dots, x_{k-1}) and an intercept. The parenthetical term in Eqn (10) is a useful index of measurement quality of x_k because it is bounded between zero and one. Eqn (10) implies that the OLS estimate of β_k is attenuated, or smaller in absolute value, than the true value. Examination of Eqn (10) also lends insight into the sources of this bias. *Ceteris paribus*, the higher the error variance relative to the variance of x_k , the greater the bias. Additionally, *ceteris paribus*, the more collinear x_k^* is with the other regressors (x_1, \dots, x_{k-1}) , the worse the attenuation bias.

Measurement error in x_k generally produces inconsistent estimates of *all* of the β_j , even when the measurement error, w , is *uncorrelated* with the other explanatory variables. This additional bias operates via the covariance matrix of the explanatory variables. The probability limit of the coefficient on a perfectly measured variable, β_j , $j \neq k$, is:

$$\text{plim} \left(\hat{\beta}_j \right) = \phi_{yx_j} - \text{plim} \left(\hat{\beta}_k \right) \phi_{xx_j}, \quad j \neq k, \quad (11)$$

where ϕ_{yx_j} is the coefficient on x_j in a population linear projection of y on (x_1, \dots, x_{k-1}) , and ϕ_{xx_j} is the coefficient on x_j in a population linear projection of x_k on (x_1, \dots, x_{k-1}) .

Eqn (11) is useful for determining the magnitude and sign of the biases in the coefficients on the perfectly measured regressors. First, if x_k^* is uncorrelated with all of the x_j , then this regressor can be left out of the regression, and the plim of the OLS estimate of β_j is ϕ_{yx_j} , which is the first term in Eqn (11). Intuitively, the measurement error in x_k cannot infect the other coefficients via correlation among the covariates if this correlation is zero. More generally, although bias in the OLS estimate of the coefficient β_k is always toward zero, bias in the other coefficients can go in either direction and can be quite large. For instance, if ϕ_{xx_j} is positive, and $\beta_k > 0$, then the OLS estimate of β_j is biased upward. As a simple numerical example, suppose, $\phi_{xx_j} = 1$, $\phi_{yx_j} = 0.2$, and the true value of $\beta_k = 0.1$. Then from Eqn (11) the true value of $\beta_j = 0.1$. However, if the biased OLS estimate of β_k is 0.05, then we can again use Eqn (11) to see that the biased OLS estimate of β_j is 0.15. If the measurement quality index in Eqn (10) is sufficiently low so that attenuation bias is severe, and if ϕ_{xx_j} is sufficiently large, then even if the true value of β_j is negative, $j \neq k$, the OLS estimate can be positive.

What if more than one variable is measured with error under the classic errors-in-variables assumption? Clearly, OLS will produce inconsistent estimates of all the parameter estimates. Unfortunately, little research on the direction and magnitude of these inconsistencies exists because biases in this case are typically unclear and complicated to derive (e.g. Klepper and Leamer, 1984). It is safe to say that bias is not necessarily toward zero and that it can be severe.

A prominent example of measurement error in corporate finance arises in regressions of investment on Tobin's q and cash flow. Starting with Fazzari, Hubbard, and Petersen (1988), researchers have argued that if a firm cannot obtain outside financing for its investment projects, then the firm's investment should be highly correlated with the availability of internal funds. This line of argument continues with the idea that if one regresses investment on a measure of investment opportunities (in this case Tobin's q) and cash flow, the coefficient on cash flow should be large and positive for groups of firms believed to be financially

constrained. The measurement error problem here is that Tobin's q is an imperfect proxy for true investment opportunities (marginal q) and that cash flow is highly positively correlated with Tobin's q . In this case, Eqn (11) shows that because this correlation, ϕ_{xx_j} , is positive, the coefficient on cash flow, β_j , is biased upwards. Therefore, even if the true coefficient on cash flow is zero, the biased OLS estimate can be positive. This conjecture is confirmed, for example, by the evidence in Erickson and Whited (2000) and Cummins, Hassett, and Oliner (2006).

2.2 Potential Outcomes and Treatment Effects

Many studies in empirical corporate finance compare the outcomes of two or more groups. For example, Sufi (2009) compares the behavior of firms before and after the introduction of bank loan ratings to understand the implications of debt certification. Faulkender and Petersen (2010) compare the behavior of firms before and after the introduction of the American Jobs Creation Act to understand the implications of tax policy. Bertrand and Mullainathan (2003) compare the behavior of firms and plants in states passing state antitakeover laws with those in states without such laws. The quantity of interest in each of these studies is the causal effect of a binary variable(s) on the outcome variables. This quantity is referred to as a treatment effect, a term derived from the statistical literature on experiments.

Much of the recent econometrics literature examining treatment effects has adopted the potential outcome notation from statistics (Rubin, 1974 and Holland, 1986). This notation emphasizes both the quantities of interest, i.e., treatment effects, and the accompanying econometric problems, i.e., endogeneity. In this subsection, we introduce the potential outcomes notation and various treatment effects of interest that we refer to below. We also show its close relation to the linear regression model (Eqn (1)). In addition to providing further insight into endogeneity problems, we hope to help researchers in empirical corporate finance digest the econometric work underlying the techniques we discuss here.

2.2.1 Notation and Framework

We begin with an observable treatment indicator, d , equal to one if treatment is received and zero otherwise. Using the examples above, treatment could correspond to the introduction of bank loan ratings, the introduction of the Jobs Creation Act, or the passage of a state antitakeover law. Observations receiving treatment are referred to as the treatment group; observations not receiving treatment are referred to as the control group. The observable

outcome variable is again denoted by y , examples of which include investment, financial policy, executive compensation, etc.

There are two potential outcomes, denoted $y(1)$ and $y(0)$, corresponding to the outcomes under treatment and control, respectively. For example, if $y(1)$ is firm investment in a state that passed an antitakeover law, then $y(0)$ is that same firm's investment in the same state had it not passed an antitakeover law. The treatment effect is the difference between the two potential outcomes, $y(1) - y(0)$.¹⁰

Assuming that the expectations exist, one can compute various average effects including:

$$\text{Average Treatment Effect (ATE)} : E[y(1) - y(0)], \quad (12)$$

$$\text{Average Treatment Effect of the Treated (ATT)} : E[y(1) - y(0)|d = 1], \quad (13)$$

$$\text{Average Treatment Effect of the Untreated (ATU)} : E[y(1) - y(0)|d = 0]. \quad (14)$$

The ATE is the expected treatment effect of a subject randomly drawn from the population. The ATT and ATU are the expected treatment effects of subjects randomly drawn from the subpopulations of treated and untreated, respectively. Empirical work tends to emphasize the first two measures and, in particular, the second one.¹¹

The notation makes the estimation problem immediately clear. For each subject in our sample, we only observe one potential outcome. The outcome that we do not observe is referred to as the counterfactual. That is, the observed outcome in the data is either $y(1)$ or $y(0)$ depending on whether the subject is treated ($d = 1$) or untreated ($d = 0$). Mathematically, the observed outcome is

$$\begin{aligned} y &= \begin{cases} y(0) & \text{if } d = 0 \\ y(1) & \text{if } d = 1 \end{cases} \\ &= y(0) + d[y(1) - y(0)] \end{aligned} \quad (15)$$

Thus, the problem of inference in this setting is tantamount to a missing data problem.

This problem necessitates the comparison of treated outcomes to untreated outcomes. To estimate the treatment effect, researchers are forced to estimate

$$E(y|d = 1) - E(y|d = 0), \quad (16)$$

¹⁰A technical assumption required for the remainder of our discussion is that the treatment of one unit has no effect on the outcome of another unit, perhaps through peer effects or general equilibrium effects. This assumption is referred to as the stable unit treatment value assumption (Angrist, Imbens, and Rubin, 1996).

¹¹Yet another quantity studied in the empirical literature is the Local Average Treatment Effect or LATE (Angrist and Imbens (1994)). This quantity will be discussed below in the context of regression discontinuity design.

or

$$E(y|d = 1, X) - E(y|d = 0, X), \quad (17)$$

if the researcher has available observable covariates $X = (x_1, \dots, x_k)$ that are relevant for y and correlated with d . Temporarily ignoring the role of covariates, Eqn (16) is just the difference in the average outcomes for the treated and untreated groups. For example, one could compute the average investment of firms in states not having passed an antitakeover law and subtract this estimate from the average investment in states that have passed an antitakeover law. The relevant question is: does this difference identify a treatment effect, such as the ATE or ATT?

Using Eqn (15), we can rewrite Eqn (16) in terms of potential outcomes

$$\begin{aligned} E(y|d = 1) - E(y|d = 0) &= \{E[y(1)|d = 1] - E[y(0)|d = 1]\} \\ &+ \{E[y(0)|d = 1] - E[y(0)|d = 0]\}. \end{aligned} \quad (18)$$

The first difference on the right hand side of Eqn (18) is the ATT (Eqn (13)). The second difference is the selection bias. Thus, a simple comparison of treatment and control group averages does not identify a treatment effect. Rather, the estimate of the ATT is confounded by a selection bias term representing nonrandom assignment to the two groups.

One solution to this selection bias is random assignment.¹² In other words, if the econometrician could let the flip of a coin determine the assignment of subjects to treatment and control groups, then a simple comparison of average outcomes would identify the causal effect of treatment. To see this, note that assignment, d , is independent of potential outcomes, $(y(0), y(1))$, under random assignment so that the selection term is equal to zero.

$$E[y(0)|d = 1] - E[y(0)|d = 0] = E[y(0)|d = 1] - E[y(0)|d = 1] = 0 \quad (19)$$

The independence allows us to change the value of the conditioning variable without affecting the expectation.

Independence also implies that the ATT is equal to the ATE and ATU since

$$\begin{aligned} E(y|d = 1) &= E[y(1)|d = 1] = E[y(1)], \text{ and} \\ E(y|d = 0) &= E[y(0)|d = 0] = E[y(0)]. \end{aligned}$$

¹²An interesting example of such assignment can be found in Hertzberg, Liberti, and Paravasini (2010) who use the random rotation of loan officers to investigate the role of moral hazard in communication.

The first equality in each line follows from the definition of y in Eqn (15). The second equality follows from the independence of treatment assignment and potential outcomes. Therefore,

$$\begin{aligned} \text{ATT} &= E[y(1)|d = 1] - E[y(0)|d = 0] \\ &= E[y(1)] - E[y(0)] \\ &= \text{ATE}. \end{aligned}$$

A similar argument shows equality with the ATU.

Intuitively, randomization makes the treatment and control groups comparable in that any observable (or unobservable) differences between the two groups are small and due to chance error. Technically, randomization ensures that our estimate of the counterfactual outcome is unbiased. That is, our estimates of what treated subjects' outcomes would have been had they not been treated — or control subjects' outcomes had they been treated—are unbiased. Thus, without random assignment, a simple comparison between the treated and untreated average outcomes is not meaningful.¹³

One may argue that, unlike regression, we have ignored the ability to control for differences between the two groups with exogenous variables, (x_1, \dots, x_k) . However, accounting for observable differences is easily accomplished in this setting by expanding the conditioning set to include these variables, as in Eqn (17). For example, the empirical problem from Eqn (18) after accounting for covariates is just

$$\begin{aligned} E(y|d = 1, X) - E(y|d = 0, X) &= \{E[y(1)|d = 1, X] - E[y(0)|d = 1, X]\} \\ &\quad + \{E[y(0)|d = 1, X] - E[y(0)|d = 0, X]\}. \end{aligned} \quad (20)$$

where $X = (x_1, \dots, x_k)$. There are a variety of ways to estimate these conditional expectations. One obvious approach is to use linear regression. Alternatively, one can use more flexible and robust nonparametric specifications, such as kernel, series, and sieve estimators. We discuss some of these approaches below in the methods sections.

Eqn (20) shows that the difference in mean outcomes among the treated and untreated, conditional on X , is still equal to the ATT plus the selection bias term. In order for this term to be equal to zero, one must argue that the treatment assignment is independent of the potential outcomes conditional on the observable control variables. In essence, controlling for observable differences leaves nothing but random variation in the treatment assignment.

¹³The weaker assumption of mean independence, as opposed to distributional independence, is all that is required for identification of the treatment effects. However, it is more useful to think in terms of random variation in the treatment assignment, which implies distributional independence.

To illustrate these concepts, we turn to an example, and then highlight the similarities and differences between treatment effects and selection bias, and linear regression and endogeneity.

2.2.2 An Example

To make these concepts concrete, consider identifying the effect of a credit rating on a firm's leverage ratio, as in Tang (2009). Treatment is the presence of credit rating so that $d = 1$ across firms with a rating and $d = 0$ for those without. The outcome variable y is a measure of leverage, such as the debt-equity ratio. For simplicity, assume that all firms are affected similarly by the presence of a credit rating so that the treatment effect is the same for firms.

A naive comparison of the average leverage ratio of rated firms to unrated firms is unlikely to identify the causal effect of credit ratings on leverage because credit ratings are not randomly assigned with respect to firms' capital structures. Eqn (18) shows the implications of this nonrandom assignment for estimation. Firms that choose to get a rating are more likely to have more debt, and therefore higher leverage, than firms that choose not to have a rating. That is, $E[y(0)|d = 1] > E[y(0)|d = 0]$, implying that the selection bias term is positive and the estimated effect of credit ratings on leverage is biased up.

Of course, one can and should control for observable differences between firms that do and do not have a credit rating. For example, firms with credit ratings tend to be larger on average, and many studies have shown a link between leverage and firm size (e.g., Titman and Wessels, 1988). Not controlling for differences in firms size would lead to a positive selection bias akin to an omitted variables bias in a regression setting. In fact, there are a number of observable differences between firms with and without a credit rating (Lemmon and Roberts, 2010), all of which should be included in the conditioning set, X .

The problem arises from *unobservable* differences between the two groups, such that the selection bias term in Eqn (20) is still nonzero. Firms' decisions to obtain credit ratings, as well as the ratings themselves, are based upon nonpublic information that is likely relevant for capital structure. Examples of this private information include unreported liabilities, corporate strategy, anticipated competitive pressures, expected revenue growth, etc. It is the relation between these unobservable measures, capital structure, and the decision to obtain a credit rating that creates the selection bias preventing researchers from estimating the quantity of interest, namely, the treatment effect of a credit rating.

What is needed to identify the causal effects of credit ratings is random or exogenous variation in their assignment. Methods for finding and exploiting such variation are discussed

below.

2.2.3 The Link to Regression and Endogeneity

We can write the observable outcome y just as we did in Eqn (1), except that there is only one explanatory variable, the treatment assignment indicator d . That is,

$$y = \beta_0 + \beta_1 d + u, \tag{21}$$

where

$$\begin{aligned} \beta_0 &= E[y(0)], \\ \beta_1 &= y(1) - y(0), \text{ and} \\ u &= y(0) - E[y(0)]. \end{aligned}$$

Plugging these definitions into Eqn (21) recovers the definition of y in terms of potential outcomes, as in Eqn (15).

Now consider the difference in conditional expectations of y , as defined in our regression Eqn (21).

$$\begin{aligned} E(y|d = 1) - E(y|d = 0) &= [\beta_0 + \beta_1 + E(u|d = 1)] - [\beta_0 + E(u|d = 0)] \\ &= \beta_1 + [E(u|d = 1) - E(u|d = 0)] \\ &= \beta_1 + [E(y(0)|d = 1) - E(y(0)|d = 0)] \end{aligned}$$

The last equality follows from the definition of u above. What this derivation shows is that unless treatment assignment is random with respect to the potential outcomes, i.e., $E(y(0)|d = 1) = E(y(0)|d = 0)$, the regression model is unidentified. OLS estimation of Eqn (21) will not recover the parameter β_1 , rather, the estimate $\hat{\beta}_1$ will be confounded by the selection bias term.

In the context of our credit rating example from above, the OLS estimate will not reveal the effect of a credit rating on leverage. The estimate will reflect the effect of a credit rating *and* the effect of any other differences between the treatment and control groups that are relevant for leverage. The bottom line is that the regression will not answer the question of interest—what is the effect of a credit rating on leverage—because the estimate is not an estimate of a quantity of interest—e.g., average treatment effect of the treated.

Of course, one can incorporate a variety of controls in Eqn (21), such as firm size, the market-to-book ratio, asset tangibility, profitability, etc. (Rajan and Zingales, 1995). Doing

so will help mitigate the selection problem but, ultimately, it will not solve the problem if the treatment and control groups differ along unobservables that are related to leverage and its determinants. Equivalently, if there is an omitted variable in u that is correlated with d or even X , the OLS estimates of *all* the parameters are likely to be inconsistent. Instead, one must find exogenous variation in credit ratings, which is the focus of Tang (2009). Thus, the implication of nonrandom assignment for estimating causal treatment effects is akin to the implications of including an endogenous dummy variable in a linear regression. As such, the solutions are similar: find random variation in the treatment assignment or, equivalently, exogenous variation in the dummy variable.

2.2.4 Heterogeneous Effects

A similar intuition holds in the context of heterogeneous treatment effects, or treatment effects that vary across subjects. To make things concrete, consider the possibility that the effect of a credit rating varies across firms. In this case, Eqn (21) becomes

$$y = \beta_0 + \tilde{\beta}_1 d + u,$$

where the treatment effect $\tilde{\beta}_1$ is now a random variable.¹⁴ The difference in expected outcomes for the treated and untreated groups is

$$\begin{aligned} E(y|d = 1) - E(y|d = 0) &= E(\tilde{\beta}_1|d = 1) + [E(u|d = 1) - E(u|d = 0)] \\ &= E(\tilde{\beta}_1|d = 1) + [E(y(0)|d = 1) - E(y(0)|d = 0)] \quad (22) \end{aligned}$$

The conditional expectation of $\tilde{\beta}_1$ is the ATT, and the difference in brackets is the selection bias term.

To recover the ATE, note that

$$\begin{aligned} E(\tilde{\beta}_1) &= Pr(d = 0)E(\tilde{\beta}_1|d = 0) + Pr(d = 1)E(\tilde{\beta}_1|d = 1) \\ &= Pr(d = 0) \left[E(\tilde{\beta}_1|d = 0) - E(\tilde{\beta}_1|d = 1) \right] + E(\tilde{\beta}_1|d = 1) \end{aligned}$$

Using this result and Eqn (22) yields

$$\begin{aligned} E(y|d = 1) - E(y|d = 0) &= E(\tilde{\beta}_1) + [E(y(0)|d = 1) - E(y(0)|d = 0)] \\ &\quad - Pr(d = 0) \left[E(\tilde{\beta}_1|d = 0) - E(\tilde{\beta}_1|d = 1) \right] \end{aligned}$$

¹⁴The sample analogue of this specification allows the treatment effect to vary across observations as such

$$y_i = \beta_0 + \beta_{1i}d_i + u_i.$$

The first term on the right hand side is the ATE, the second term the selection bias. With heterogeneous effects, there is an additional bias term corresponding to the difference in expected gains from the treatment across treatment and control groups. Of course, when the treatment assignment is randomized, both of these bias terms equal zero so that the difference in means recovers the ATE.

2.3 Identifying and Discussing the Endogeneity Problem

Before discussing how to address endogeneity problems, we want to emphasize a more practical matter. A necessary first step in any empirical corporate finance study focused on disentangling alternative hypotheses or identifying causal effects is identifying the endogeneity problem and its implications for inference. Unsurprisingly, it is difficult, if not impossible, to address a problem without first understanding it. As such, we encourage researchers to discuss the primary endogeneity concern in their study.

There are a number of questions that should be answered before putting forth a solution. Specifically, what is the endogenous variable(s)? Why are they endogenous? What are the implications for inferences of the endogeneity problems? In other words, what are the alternative hypotheses about which one should be concerned? Only after answering these questions can researchers put forth a solution to the endogeneity problem.

3. Instrumental Variables

In this section we discuss instrumental variables (IV) as a way to deal with endogeneity, with an emphasis on the hurdles and challenges that arise when trying to implement IV in corporate finance settings. We first outline the basic econometric framework and discuss how to find instruments. We then move to the issue of weak instruments and to the tradeoff between internal and external validity that one naturally encounters in IV estimation.

3.1 What are Valid Instruments?

We start this section with the single equation linear model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u. \quad (23)$$

We assume that key assumptions 1 through 3 for the consistency of the OLS estimator hold (see section 2.1). However, we relax the assumption that $\text{cov}(x_j, u) = 0, \forall j$ and, for

simplicity, consider the case in which one regressor, x_k , is correlated with u . In this case all of the regression coefficients are biased except in the special and unlikely case that x_k is uncorrelated with the rest of the regressors. In this particular case, only the estimate of β_k is biased.

The standard remedy for endogeneity is finding an instrument for the endogenous regressor, x_k . An instrument, z , is a variable that satisfies two conditions that we refer to as the relevance and exclusion conditions. The first condition requires that the partial correlation between the instrument and the endogenous variable not be zero. In other words, the relevance condition requires that the coefficient γ in the regression

$$x_k = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + \gamma z + v \quad (24)$$

does not equal zero. This condition is *not* equivalent to nonzero correlation between x_k and z . It refers to the correlation between x_k and z after netting out the effects of all other exogenous variables. Fortunately, this condition is empirically testable. Estimate Eqn (24) via OLS and test the null hypothesis that $\gamma = 0$ against the alternative that $\gamma \neq 0$. However, as we discuss below, the usual t-test in this instance may be inappropriate.

The exclusion condition requires that $\text{cov}(z, u) = 0$. The name for this condition derives from the exclusion of the instrument from Eqn (23). In conjunction with the relevance condition, the exclusion restriction implies that the only role that the instrument z plays in influencing the outcome y is through its affect on the endogenous variable x_k . Together with the relevance condition, the exclusion condition identifies the parameters in Eqn (23).¹⁵ However, unlike the relevance condition, the exclusion condition cannot be tested because the regression error term, u , is unobservable.

There is nothing restricting the number of instruments to just one. Any variable satisfying both relevance and exclusion conditions is a valid instrument. In the case that there are multiple instruments $z = (z_1, \dots, z_m)$, the relevance condition can be tested with a test of the joint null hypothesis that $\gamma_1 = 0, \dots, \gamma_m = 0$ against the alternative hypothesis that at least one γ coefficient is nonzero in the model

$$x_k = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + \gamma_1 z_1 + \cdots + \gamma_m z_m + v. \quad (25)$$

The exclusion restriction requires the correlation between each instrument and the error term u in Eqn (23) to be zero (i.e., $\text{cov}(z_j, u) = 0$ for $j = 1, \dots, m$).

¹⁵Write Eqn (1) as $y = XB + u$, where $B = (\beta_0, \beta_1, \dots, \beta_k)'$ and $X = (1, x_1, \dots, x_k)$. Let $Z = (1, x_1, \dots, x_{k-1}, z)$ be the vector of all exogenous variables. Premultiply the vector equation by Z' and take expectations so that $E(Z'y) = E(Z'X)B$. Solving for B yields $B = E(Z'X)^{-1}E(Z'y)$. In order for this equation to have a unique solution, assumptions 3 and 4 (or 4a) must hold.

Likewise, there is nothing restricting the number of endogenous variables to just one. Consider the model,

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \beta_{k+1} x_{k+1} + \dots + \beta_{k+h-1} x_{k+h-1} + u, \quad (26)$$

where (x_1, \dots, x_{k-1}) are the $k - 1$ exogenous regressors and (x_k, \dots, x_{k+h-1}) are the h endogenous regressors. In this case, we must have at least as many instruments (z_1, \dots, z_m) as endogenous regressors in order for the coefficients to be identified, i.e., $m \geq h$. The exclusion restriction is unchanged from the previous paragraph: all instruments must be uncorrelated with the error term u . The relevance condition is similar in spirit except now there is a system of relevance conditions corresponding to the system of endogenous variables.

$$\begin{aligned} x_k &= \alpha_{10} + \alpha_{11} x_1 + \dots + \alpha_{1k-1} x_{k-1} + \gamma_{11} z_1 + \dots + \gamma_{1m} z_m + v. \\ &\vdots \\ x_{k+h-1} &= \alpha_{h0} + \alpha_{h1} x_1 + \dots + \alpha_{hk-1} x_{k-1} + \gamma_{h1} z_1 + \dots + \gamma_{hm} z_m + v. \end{aligned}$$

The relevance condition in this setting is analogous to the relevance condition in the single-instrument case: the instruments must be “fully correlated” with the regressors. Formally, $E(Xz)$ has to be of full column rank, that is, $\text{rank}(Xz) = k$.

Models with more instruments (m) than endogenous variables (h) are said to be overidentified and there are $(m - h)$ overidentifying restrictions. For example, with only one endogenous variable, we need only one valid instrument to identify the coefficients (see footnote 15). Hence, the additional instruments are unnecessary from an identification perspective. What is the optimal number of instruments? From an asymptotic efficiency perspective, more instruments is better. However, from a finite sample perspective, more instruments is not necessarily better and can even exacerbate the bias inherent in 2SLS.¹⁶

3.2 Estimation

Given a set of instruments, the question is how to use them to consistently estimate the parameters in Eqn (23). The most common approach is two-stage least squares (2SLS). As the name suggests, 2SLS can conceptually be broken down into two parts.

1. Estimate the predicted values, \hat{x}_k , by regressing the endogenous variable x_k on all of the exogenous variables—controls (x_1, \dots, x_{k-1}) and instruments (z_1, \dots, z_m) —as in

¹⁶Although instrumental variables methods such as 2SLS produce consistent parameter estimates, they do not produce unbiased parameter estimates when at least one explanatory variable is endogenous.

Eqn (24). (One should also test the significance of the instruments in this regression to ensure that the relevance condition is satisfied.)

2. Replace the endogenous variable x_k with its predicted values from the first stage \hat{x}_k , and regress the outcome variable y on all of the control variables (x_1, \dots, x_{k-1}) and \hat{x}_k .

This two-step procedure can be done all at once. Most software programs do exactly this, which is useful because the OLS standard errors in the second stage are incorrect.¹⁷ However, thinking about the first and second stages separately is useful because doing so underscores the intuition that variation in the endogenous regressor x_k has two parts: the part that is uncorrelated with the error (“good” variation) and the part that is correlated with the error (“bad” variation). The basic idea behind IV regression is to isolate the “good” variation and disregard the “bad” variation.

3.3 Where do Valid Instruments Come From? Some Examples

Good instruments can come from biological or physical events or features. They can also sometimes come from institutional changes, as long as the economic question under study was not one of the reasons for the institutional change in the first place. The only way to find a good instrument is to understand the economics of the question at hand. The question one should always ask of a potential instrument is, “Does the instrument affect the outcome only via its effect on the endogenous regressor?” To answer this question, it is also useful to ask whether the instrument is likely to have any effect on the dependent variable—either the observed part (y) or the unobserved part (u). If the answer is yes, the instrument is probably not valid.¹⁸

A good example of instrument choice is in Bannedsen et al. (2007), who study CEO succession in family firms. They ask whether replacing an outgoing CEO with a family member hurts firm performance. In this example, performance is the dependent variable, y , and family CEO succession is the endogenous explanatory variable, x_k . The characteristics of the firm and family that cause it to choose a family CEO may also cause the change in performance. In other words, it is possible that an omitted variable causes both y and x_k , thereby leading to a correlation between x_k and u . In particular, a nonfamily CEO might

¹⁷The problem arises from the use of a generated regressor, \hat{x}_k , in the second stage. Because this regressor is itself an estimate, it includes estimation error. This estimation error must be taken into account when computing the standard error of its, and the other explanatory variables’, coefficients.

¹⁸We refer the reader to the paper by Conley, Hansen, and Rossi (2010) for an empirical approach designed to address imperfect instruments.

be chosen to “save” a failing firm, and a family CEO might be chosen if the firm is doing well or if the CEO is irrelevant for firm performance. This particular example is instructive because the endogeneity—the correlation between the error and regressor—is directly linked to specific economic forces. In general, good IV studies always point out specific sources of endogeneity and link these sources directly to the signs and magnitudes (if possible) of regression coefficients.

Bennedsen et al. (2007) choose an instrumental variables approach to isolate exogenous variation in the CEO succession decision. Family characteristics such as size and marital history are possible candidates, because they are highly correlated with the decision to appoint a family CEO. However, if family characteristics are in part an outcome of economic incentives, they may not be exogenous. That is, they may be correlated with firm performance. The instrument, z , Bennedsen et al. (2007) choose is the gender of the first-born child of a departing CEO. On an intuitive level, this type of biological event is unlikely to affect firm performance, and Bennedsen, et al. document that boy-first firms are similar to girl-first firms in terms of a variety of measures of performance. Although not a formal test of the exclusion restriction, this type of informal check is always a useful and important part of any IV study.

The authors then show that CEOs with boy-first families are significantly more likely to appoint a family CEO in their first stage regressions, i.e., the relevance conditional is satisfied. In their second stage regressions, they find that the IV estimates of the negative effect of in-family CEO succession are much larger than the OLS estimates. This difference is exactly what one would expect if outside CEOs are likely to be appointed when firms are doing poorly. By instrumenting with the gender of the first-born, Bennedsen et al. (2007) are able to isolate the exogenous or random variation in family CEO succession decisions. And, in doing so, readers can be confident that they have isolated the causal effect of family succession decisions on firm performance.¹⁹

3.4 So Called Tests of Instrument Validity

As mentioned above, it is impossible to test directly the assumption that $\text{cov}(z, u) = 0$ because the error term is unobservable. Instead, researchers must defend this assumption in two ways. First, compelling arguments relying on economic theory and a deep understanding of the relevant institutional details are the most important elements of justifying an instrument’s validity. Second, a number of falsification tests to rule out alternative hypotheses

¹⁹Other examples of instrumental variables applications in corporate finance include: Guiso, Sapienza, and Zingales (2004), Becker (2007), Giroud et al. (2010), and Chaney, Sraer, and Thesmar (in press).

associated with endogeneity problems can also be useful. For example, consider the evidence put forth by Bannedsen et al. (2007) showing that the performance of firms run by CEOs with a first born boy is no different from that of firms run by CEOs with a first born girl.

In addition, a number of statistical specification tests have been proposed. The most common one in an IV setting is a test of the overidentifying restrictions of the model, assuming one can find more instruments than endogenous regressors. On an intuitive level, the test of overidentifying restrictions tests whether all possible subsets of instruments that provide exact identification provide the same estimates. In the population, these different subsets should produce identical estimates if the instruments are all truly exogenous.

Unfortunately, this test is unlikely to be useful for three reasons. First, the test assumes that at least one instrument is valid, yet which instrument is valid and why is left unspecified. Further, in light of the positive association between finite sample bias and the number of instruments, if a researcher has one good instrument the choice to find more instruments is not obvious. Second, finding instruments in corporate finance is sufficiently difficult that it is rare for a researcher to find several. Third, although the overidentifying test can constitute a useful diagnostic, it does not always provide a good indicator of model misspecification. For example, suppose we expand the list of instruments that are uncorrelated with u . We will not raise the value of the test statistic, but we will increase the degrees of freedom used to construct the regions of rejection. This increase artificially raises the critical value of the chi-squared statistic and makes rejection less likely. In short, these tests may lack power.

Ultimately, good instruments are both rare and hard to find. There is no way to test their validity beyond rigorous economic arguments and, perhaps, a battery of falsification tests designed to rule out alternative hypotheses. As such, we recommend thinking carefully about the economic justification—either via a formal model or rigorous arguments—for the use of a particular instrument.

3.5 The Problem of Weak Instruments

The last two decades have seen the development of a rich literature on the consequences of weak instruments. As surveyed in Stock, Wright, and Yogo (2002), instruments that are weakly correlated with the endogenous regressors can lead to coefficient bias in finite samples, as well as test statistics whose finite sample distributions deviate sharply from their asymptotic distributions. This problem arises naturally because those characteristics, such as randomness, that make an instrument a source of exogenous variation may also make the instrument weak.

The bias arising from weak instruments can be severe. To illustrate this issue, we consider a case in which the number of instruments is larger than the number of endogenous regressors. In this case Hahn and Hausman (2005) show that the finite-sample bias of two-stage least squares is approximately

$$\frac{j\rho(1-r^2)}{nr^2}, \quad (27)$$

where j is the number of instruments, ρ is the correlation coefficient between x_k and u , n is the sample size, and r^2 is the R^2 of the first-stage regression. Because the r^2 term is in the denominator of Eqn (27), even with a large sample size, this bias can be large.

A number of diagnostics have been developed in order to detect the weak instruments problem. The most obvious clue for extremely weak instruments is large standard errors because the variance of an IV estimator depends inversely on the covariance between the instrument and the exogenous variable. However, in less extreme cases weak instruments can cause bias and misleading inferences even when standard errors are small.

Stock and Yogo (2005) develop a diagnostic based on the Cragg and Donald F statistic for an underidentified model. The intuition is that if the F statistic is low, the instruments are only weakly correlated with the endogenous regressor. They consider two types of null hypotheses. The first is that the bias of two-stage least squares is less than a given fraction of the bias of OLS, and the second is that the actual size of a nominal 5% two-stage least squares t-test is no more than 15%. The first null is useful for researchers that are concerned about bias, and the second is for researchers concerned about hypothesis testing.

They then tabulate critical values for the F statistic that depend on the given null. For example, in the case when the null is that the two-stage least squares bias is less than 10% of the OLS bias, when the number of instruments is 3, 5, and 10, the suggested critical F-values are 9.08, 10.83, and 11.49, respectively. The fact that the critical values increase with the number of instruments implies that adding additional low quality instruments is not the solution to a weak-instrument problem. As a practical matter, in any IV study, it is important to report the first stage regression, including the R^2 . For example, Bannedsen et al. (2007) report that the R^2 of their first stage regression (with the instrument as the only explanatory variable) is over 40%, which indicates a strong instrument. They confirm this strength with subsequent tests of the relevance condition using the Stock and Yogo (2005) critical values.²⁰

²⁰Hahn and Hausman (2005) propose a test for weak instruments in which the null is that the instruments are strong and the alternative is that the instruments are weak. They make the observation that under the null the choice of the dependent variable in Eqn (23) should not matter in an IV regression. In other words, if the instruments are strong, the IV estimates from Eqn (23) should be asymptotically the same as the IV

Not only do weak instruments cause bias, but they distort inference. Although a great deal of work has been done to develop tests that are robust to the problem of weak instruments, much of this work has been motivated by macroeconomic applications in which data are relatively scarce and in which researchers are forced to deal with whatever weak instruments they have. In contrast, in a data rich field like corporate finance, we recommend spending effort in finding strong—and obviously valid—instruments rather than in dealing with weak instruments.

3.6 Lagged Instruments

The use of lagged dependent variables and lagged endogenous variables has become widespread in corporate finance.²¹ The original economic motivation for using dynamic panel techniques in corporate finance comes from estimation of investment Euler equations using firm-level panel data (Whited, 1992, Bond and Meghir, 1994). Intuitively, an investment Euler equation can be derived from a perturbation argument that states that the marginal cost of investing today is equal, at an optimum, to the expected discounted cost of delaying investment until tomorrow. This latter cost includes the opportunity cost of the foregone marginal product of capital as well as any direct costs.

Hansen and Singleton (1982) point out that estimating any Euler equation—be it for investment, consumption, inventory accumulation, labor supply, or any other intertemporal decision—requires an assumption of rational expectations. This assumption allows the empirical researcher to replace the expected cost of delaying investment, which is inherently unobservable, with the actual cost plus an expectational error. The intuition behind this replacement is straightforward: as a general rule, what happens is equal to what one expects plus one’s mistake. Further, the mistake has to be orthogonal to any information available at the time that the expectation was made; otherwise, the expectation would have been different. This last observation allows lagged endogenous variables to be used as instruments to estimate the Euler equation.

It is worth noting that the use of lagged instruments in this case is motivated by the characterization of the regression error as an expectational error. Under the joint null hypothesis that the model is correct and that agents have rational expectations, lagged instruments can be argued to affect the dependent variable only via their effect on the endogenous regressors.

estimates of a regression in which y and x_k have been swapped. Their test statistic is then based on this equality.

²¹For example, see Flannery and Rangan (2006), Huang and Ritter (2009), and Iliev and Welch (2010) for applications and analysis of dynamic panel data models in corporate capital structure.

This intuition does not carry over to a garden variety regression. We illustrate this point in the context of a standard capital structure regression from Rajan and Zingales (1995), in which the book leverage ratio, y_{it} , is the dependent variable and in which the regressors are the log of sales, s_{it} , the market-to-book ratio, m_{it} , the lagged ratio of operating income to assets, o_{it} , and a measure of asset tangibility, k_{it} :

$$y_{it} = \beta_0 + \beta_1 s_{it} + \beta_2 m_{it} + \beta_3 o_{it} + \beta_4 k_{it} + u_{it}.$$

These variables are all determined endogenously as the result of an explicit or implicit managerial optimization, so simultaneity might be a problem. Further, omitted variables are also likely a problem since managers rely on information unavailable to econometricians but likely correlated with the included regressors. Using lagged values of the dependent variable and endogenous regressors as instruments requires one to believe that they affect leverage only via their correlation with the endogenous regressors. In this case, and in many others in corporate finance, this type of argument is hard to justify. The reason here is that all five of these variables are quite persistent. Therefore, if current operating income is correlated with u_{it} , then lagged operating income is also likely correlated with u_{it} . Put differently, if a lagged variable is correlated with the observed portion of leverage, then it is hard to argue that it is uncorrelated with the unobserved portion, that is, u_{it} .

In general, we recommend thinking carefully about the economic justification for using lagged instruments. To our knowledge, no such justification has been put forth in corporate finance outside the Euler equation estimation literature. Rather, valid instruments for determinants of corporate behavior are more likely to come from institutional changes and nonfinancial variables.

3.7 Limitations of Instrumental Variables

Unfortunately, it is often the case that in corporate finance more than one regressor is endogenous. In this case, inference about all of the regression coefficients can be compromised if one can find instruments for only a subset of the endogenous variables. For example, suppose in Eqn (23) that both x_k and x_{k-1} are endogenous. Then even if one has an instrument z for x_k , unless z is uncorrelated with x_{k-1} , the estimate of β_{k-1} will be biased. Further, if the estimate of β_{k-1} is biased, then unless x_{k-1} is uncorrelated with the other regressors, the rest of the regression coefficients will also be biased. Thus, the burden on instruments in corporate finance is particularly steep because few explanatory variables are truly exogenous.

Another common mistake in the implementation of IV estimators is more careful attention to the relevance of the instruments than to their validity. This problem touches even the best IV papers. As pointed out in Heckman (1997), when the effects of the regressors on the dependent variable are heterogeneous in the population, even purely random instruments may not be valid. For example, in Bennedsen et al. (2007) it is possible that families with eldest daughters may still choose to have the daughter succeed as CEO of the firm if the daughter is exceptionally talented. Thus, while family CEO succession hurts firm performance in boy-first families, the option of family CEO succession in girl-first families actually improves performance. This contrast causes the IV estimator to exaggerate the negative effect of CEO succession on firm performance.

This discussion illustrates the point that truly exogenous instruments are extremely difficult to find. If even random instruments can be endogenous, then this problem is likely to be magnified with the usual non-random instruments found in many corporate finance studies. Indeed, many papers in corporate finance discuss only the relevance of the instrument and ignore any exclusion restrictions.

A final limitation of IV is that it—like all other strategies discussed in this study—faces a tradeoff between external and internal validity. IV parameter estimates are based only on the variation in the endogenous variable that is correlated with the instrument. Bennedsen et al. (2007) provide a good illustration of this issue because their instrument is binary. Their results are applicable only to those observations in which a boy-first family picks a family CEO or in which a girl-first family picks a non-family CEO. This limitation brings up the following concrete and important question. What if the family CEOs that gain succession and that are affected by primogeniture are of worse quality than the family CEOs that gain succession and that are not affected by primogeniture? Then the result of a strong negative effect of family succession is not applicable to the entire sample.

To address this point, it is necessary to identify those families that are affected by the instrument. Clearly, they are those observations that are associated with a small residual in the first stage regression. Bennedsen et al. (2007) then compare CEO characteristics across observations with large residuals (not affected by the instrument) and those with small residuals (affected by the instrument), and they find that these two groups are largely similar. In general, it is a good idea to conduct this sort of exercise to determine the external validity of IV results.

4. Difference-in-Differences Estimators

Difference-in-Differences (DD) estimators are used to recover the treatment effects stemming from sharp changes in the economic environment, government policy, or institutional environment. These estimators usually go hand in hand with the natural or quasi-experiments created by these sharp changes. However, the exogenous variation created by natural experiments is much broader than any one estimation technique. Indeed, natural experiments have been used to identify instrumental variables for 2SLS estimation and discontinuities for regression discontinuity designs discussed below.²²

The goal of this section is to introduce readers to the appropriate application of the DD estimator. We begin by discussing single difference estimators to highlight their shortcomings and to motivate DD estimators, which can overcome these shortcomings. We then discuss how one can check the internal validity of the DD estimator, as well as several extensions.

4.1 Single Cross-Sectional Differences After Treatment

One approach to estimating a parameter that summarizes the treatment effect is to compare the post-treatment outcomes of the treatment and control groups. This method is often used when there is no data available on pre-treatment outcomes. For example, Garvey and Hanka (1999) estimate the effect of state antitakeover laws on leverage by examining one year of data after the law passage. They then compare the leverage ratios of firms in states that passed the law (the treatment group) and did not pass the law (the control group). This comparison can be accomplished with a cross-sectional regression:

$$y = \beta_0 + \beta_1 d + u, \tag{28}$$

where y is leverage, and d is the treatment assignment indicator equal to one if the firm is incorporated in a state that passed the antitakeover law and zero otherwise. The difference between treatment and control group averages is β_1 .

If there are observations for several post-treatment periods, one can collapse each subject's time series of observation to one value by averaging. Eqn (28) can then be estimated using the cross-section of subject averages. This approach addresses concerns over dependence of observations within subjects (Bertrand, Duflo, and Mullainathan, 2004). Alternatively, one can modify Eqn (28) to allow the treatment effect to vary over time by interacting

²²Examples of natural experiments beyond those discussed below include Schnabl (2010), who uses the 1998 Russian default as a natural experiment to identify the transmission and impact of liquidity shocks to financial institutions.

the assignment indicator with period dummies as such,

$$y = \beta_0 + \beta_1 d \times p_1 + \cdots + \beta_T d \times p_T + u. \quad (29)$$

Here, $(\beta_1, \dots, \beta_T)$, correspond to the period-by-period differences between treatment and control groups.

From section 2, OLS estimation of Eqns (28) and (29) recovers the causal effect of the law change if and only if d is mean independent of u . Focusing on Eqn (28) and taking conditional expectations yields the familiar expression

$$\begin{aligned} E(y|d = 1) - E(y|d = 0) &= \beta_1 + [E(u|d = 1) - E(u|d = 0)] \\ &= \beta_1 + [E(y(0)|d = 1) - E(y(0)|d = 0)], \end{aligned}$$

where the second equality follows from Eqn (21). If there are any permanent unobserved differences between the treatment and control groups prior to the onset of treatment, then the selection bias is nonzero and OLS will not recover the causal effect of the law change. In the antitakeover law example, one must argue that firms are similar, prior to the passage of the laws, with regard to leverage related characteristics in states that did and did not pass the law. The validity of this argument depends crucially on why the law was changed when it was in some states and not in others.

For example, if the law was enacted to protect profitable firms from hostile raiders, then the bias term is likely to be negative. Many studies have shown a negative link between profitability and leverage (e.g., Rajan and Zingales, 1995; Frank and Goyal, 2009) implying that firms from states enacting the law ($d = 1$) tend to have lower leverage because they are more profitable. Of course, one could control for profitability, as well as a host of other variables. However, one should *not* control for variables that themselves may be affected by the treatment (e.g., other outcome variables such as investment or dividend policy). This restriction implies that most control variables should be measured prior to the onset of treatment or, in this example, passage of the law. Despite the inclusion of many observable variables, one must ensure that there are no unobservable differences related to leverage and the passage of the law that may taint inferences.

4.2 Single Time-Series Difference Before and After Treatment

A second way to estimate the treatment effect is to compare the outcome after the onset of the treatment with the outcome before the onset of treatment for just those subjects that are treated. This is a more commonly used approach in corporate finance where the

event affects all observable subjects as opposed to just a subset of subjects. For example, Bertrand, Schoar, and Thesmar (2007) examine the impact of deregulation of the French banking industry on the behavior of borrowers and banks.²³ Graham, Michaely, and Roberts (2003) compare ex-dividend day stock returns for NYSE listed firms during three different price quote eras: 1/16s, 1/8s, and decimals. Blanchard, Lopez-de-Silanes, and Shleifer (1994) compare a variety of corporate outcomes (e.g., investment, dividends, net debt issuance, net equity issuance, asset sales) before and after large legal awards. Khwaja and Mian (2008) use the unanticipated nuclear tests in Pakistan to understand the role that liquidity shocks to banks play in their lending behavior. In each case, the control and treatment groups are defined by the before- and after-treatment periods, respectively, and consist of the same subjects.

With only two time-series observations per subject—one before and one after treatment—comparisons can be accomplished with a two period panel regression using only subjects affected by the event,

$$y = \beta_0 + \beta_1 p + u, \tag{30}$$

where p equals one if the observation is made after treatment onset (i.e., post-treatment) and zero otherwise (i.e., pre-treatment). The treatment effect is given by β_1 . Alternatively, one can estimate a cross-sectional regression of first differences,

$$\Delta y = \beta_0 + \Delta u \tag{31}$$

in which case the treatment effect is given by the intercept β_0 .

With more than one period before and after the event, several options are available. First, one can estimate a level regression assuming a constant treatment effect across all post-treatment periods

$$y = \beta_0 + \beta_1 p + u,$$

where p equals one for all post-treatment periods. The differenced version of this specification is simply

$$\Delta y = \beta_0 + \beta_1 \Delta p + \Delta u,$$

though one must take care to account for the effects of differencing on the statistical properties of the error term.

²³Other examples include Sufi (2009) who examines the introduction of loan ratings.

Alternatively, one can include a dummy for each period—pre- and post-treatment—except for one:

$$y = \beta_0 + \beta_1 p_1 + \cdots + \beta_{T-1} p_{T-1} + u,$$

where p_s , $s = 1, \dots, T-1$ equals one in the s^{th} period and zero otherwise. With the estimated coefficients $(\beta_1, \dots, \beta_{T-1})$ one can visually inspect or test whether a break occurs around the date of the natural experiment, in our example the change in law. However, if the response to the law is gradual or delayed, this strategy will not work well.

Taking conditional expectations of Eqn (30) yields

$$\begin{aligned} E(y|p = 1) - E(y|p = 0) &= \beta_1 + [E(u|p = 1) - E(u|p = 0)] \\ &= \beta_1 + [E(y(0)|p = 1) - E(y(0)|p = 0)] \end{aligned}$$

The selection bias is nonzero when there exist trends in the outcome variable that are due to forces other than the treatment. Using the previous example, imagine focusing on firms in states that have passed an antitakeover law and comparing their capital structures before and after the passage. The endogeneity concern is that these firms' leverage ratios would have changed over the period of observation even if the laws had not been passed. For example, empirical evidence suggests that there is a strong counter-cyclical component to leverage ratios (Korajczyk and Levy, 2003).

This phenomenon is just another form of omitted (or mismeasured) variables bias, that stems from the inability to control perfectly for business cycle forces, financial innovation, variation in investor demand, etc. As with the cross-sectional estimator, one can incorporate controls subject to the restriction that these controls are unaffected by the treatment. However, it is no easier, or more credible, to account for all of the potentially omitted and mismeasured determinants in this time-series setting than it is in the previous cross-sectional setting. For example, throwing in the kitchen sink of macroeconomic factors is insufficient, especially if firms' sensitivities to these factors are heterogeneous. The consequence is that the OLS estimate of β_1 is biased because the regression cannot disentangle the effects of the event from all of the forces causing time series changes in the outcome variables.

An alternative strategy to addressing secular trends is to examine the outcomes for similar groups that did not receive the treatment but would be subject to similar influence from the trending variables. While one would expect to see a sharp change in behavior among the treatment group following application of the treatment, we would not expect to see such a change among groups not receiving the treatment. In fact, this approach leads us to the difference-in-differences estimator.

4.3 Double Difference Estimator: Difference-in-Differences (DD)

The two single difference estimators complement one another. The cross-sectional comparison avoids the problem of omitted trends by comparing two groups over the same time period. The time series comparison avoids the problem of unobserved differences between two different groups of firms by looking at the same firms before and after the change. The double difference, difference-in-differences (DD), estimator combines these two estimators to take advantage of both estimators' strengths.

Consider a firm-year panel dataset in which there are two time periods, one before and one after the onset of treatment, and only some of the subjects are treated. For example, Arizona passed antitakeover legislation in 1987 at which time Connecticut had not passed similar legislation (Bertrand and Mullainathan, 2003). Assuming for simplicity that the legislation was passed at the beginning of the year, 1986 represents the pre-treatment period, 1987 the post-treatment period. Firms registered in Arizona represent the treatment group, those registered in Connecticut the control group.

The regression model in levels for the DD estimator is:

$$y = \beta_0 + \beta_1 d \times p + \beta_2 d + \beta_3 p + u, \quad (32)$$

or, in differences,

$$\Delta y = \beta_0 + \beta_1 d + \Delta u, \quad (33)$$

where d is the treatment assignment variable equal to one if a firm is registered in Arizona, zero if registered in Connecticut, p is the post-treatment indicator equal to one in 1987 and zero in 1986. Intuitively, including the level d controls for permanent differences between the treatment and control groups. For example, if firms in Arizona are, on average, less levered than those in Connecticut, perhaps because they are more profitable, then β_2 should capture this variation. Likewise, including the level p controls for trends common to both treatment and control groups. For example, if leverage is increasing between 1986 and 1987 because of a secular decline in the cost of debt, β_3 will capture this variation. The variation that remains is the change in leverage experienced by firms in Arizona relative to the change in leverage experienced by firms in Connecticut. This variation is captured by β_1 , the DD estimate.

The DD estimator can also be obtained using differences of variables, as opposed to levels. With two periods, one pre- and one post-treatment, a cross-sectional regression of the change in outcomes, Δy , on a treatment group indicator variable d and the change in

control variables $\Delta x_1, \dots, \Delta x_k$, if any, will recover the treatment effect, β_1 . Mathematically, the specification is a generalization of Eqn (33):

$$\Delta y = \beta_0 + \beta_1 d + \Delta X \psi + \Delta u,$$

where ΔX is a $1 \times k$ vector of changes in the covariates, $(\Delta x_1, \dots, \Delta x_k)$.

More formally, consider the conditional expectations corresponding to the four combinations of values for the indicator variables in Eqn (32):

$$E(y|d = 1, p = 1) = \beta_0 + \beta_1 + \beta_2 + \beta_3,$$

$$E(y|d = 1, p = 0) = \beta_0 + \beta_2,$$

$$E(y|d = 0, p = 1) = \beta_0 + \beta_3,$$

$$E(y|d = 0, p = 0) = \beta_0,$$

assuming $E(u|d, p) = 0$. Arranging these conditional means into a two-by-two table and computing row and column differences produces:

Table 1: **Conditional Mean Estimates from the DD Regression Model**

	Post-Treatment	Pre-Treatment	Difference
Treatment	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_0 + \beta_2$	$\beta_1 + \beta_3$
Control	$\beta_0 + \beta_3$	β_0	β_3
Difference	$\beta_1 + \beta_2$	β_2	β_1

The inner cells of the table are the conditional means. For example, the average y for firms in the treatment group during the post-treatment period is $(\beta_0 + \beta_1 + \beta_2 + \beta_3)$. Likewise, the average y for firms in the control group during the pre-treatment period is β_0 . The outer cells correspond to differences in these conditional means. The average difference in y between the treatment and control groups during the pre-treatment period is $(\beta_0 + \beta_2) - \beta_0 = \beta_2$.

The cell in the bottom right corner is the DD estimate, which can be obtained by either (1) differencing down the rightmost column, or (2) differencing across the bottom row. These two cases can be expressed as

$$\begin{aligned} \beta_1 &= (E(y|d = 1, p = 1) - E(y|d = 1, p = 0)) - (E(y|d = 0, p = 1) - E(y|d = 0, p = 0)) \\ &= (E(y|d = 1, p = 1) - E(y|d = 0, p = 1)) - (E(y|d = 1, p = 0) - E(y|d = 0, p = 0)) \end{aligned}$$

The top line is the difference in the treatment group from the pre to post era minus the difference in the control group from the pre to post era, and the second line gives the

difference between the treatment and control group in the post era minus the difference between the treatment and control group in the pre era. Linearity ensures that these two approaches generate the same result.

4.3.1 Revisiting the Single Difference Estimators

Reconsider the single difference cross-sectional estimator of section 4.1. According to Eqn (32), this estimator can be written as

$$\begin{aligned} E(y|d = 1, p = 1) - E(y|d = 0, p = 1) &= (\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_3) \\ &= \beta_1 + \beta_2 \end{aligned}$$

This result shows that the cross-sectional difference estimator is an unbiased estimate of the treatment effect, β_1 , only if $\beta_2 = 0$. In other words, the treatment and control groups cannot differ in ways that are relevant for the outcome variable, precisely as noted above.

Now reconsider the single difference time-series estimator of section 4.2.2. According to Eqn (32), this estimator can be written as

$$\begin{aligned} E(y|d = 1, p = 1) - E(y|d = 1, p = 0) &= (\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_2) \\ &= \beta_1 + \beta_3 \end{aligned}$$

This result shows that the time-series difference estimator is an unbiased estimate of the treatment effect, β_1 , only if $\beta_3 = 0$. In other words, there can be no trends relevant for the outcome variable, precisely as noted above.

The DD estimator takes care of these two threats to the identification of the treatment effect. First, any permanent, i.e., time-invariant, difference between the treatment and control groups is differenced away by inclusion of the d indicator variable. Second, any common trend affecting both the treatment and control group is also differenced away by inclusion of the p indicator variable. These points are important, and often misunderstood: threats to the internal validity of the DD estimator cannot come from either permanent differences between the treatment and control groups, or shared trends.

4.3.2 Model Extensions

One can easily incorporate covariates, X , into Eqn (32)

$$y = \beta_0 + \beta_1 d \times p + \beta_2 d + \beta_3 p + X\Gamma + u,$$

or Eqn (33)

$$\Delta y = \beta_0 + \beta_1 d + \Delta X \Gamma + \Delta u,$$

where $X = (x_1, \dots, x_k)$ and $\Gamma = (\gamma_1, \dots, \gamma_k)'$. In each case, β_1 is the DD estimator with roughly the same interpretation as the case without covariates.

Reasons for including covariates include: efficiency, checks for randomization, and adjusting for conditional randomization. Assuming random or exogenous assignment to treatment and control groups, the OLS estimate of the treatment effect β_1 is more efficient with additional exogenous controls because these controls reduce the error variance.

If assignment is random, then including additional covariates should have a negligible effect on the estimated treatment effect. Thus, a large discrepancy between the treatment effect estimates with and without additional controls raises a red flag. If assignment to treatment and control groups is not random but dictated by an observable rule, then controlling for this rule via covariates in the regression satisfies the conditional mean zero assumption required for unbiased estimates. Regardless of the motivation, it is crucial to remember that any covariates included as controls must be unaffected by the treatment, a condition that eliminates other outcome variables and restricts most covariates to pre-treatment values.

Bertrand, Duflo, and Mullainathan (2004) propose a general model to handle multiple time periods and multiple treatment groups. The model they consider is

$$y_{igt} = \beta_1 d_{gt} + X_{igt} \Gamma + p_t + m_g + u_{igt},$$

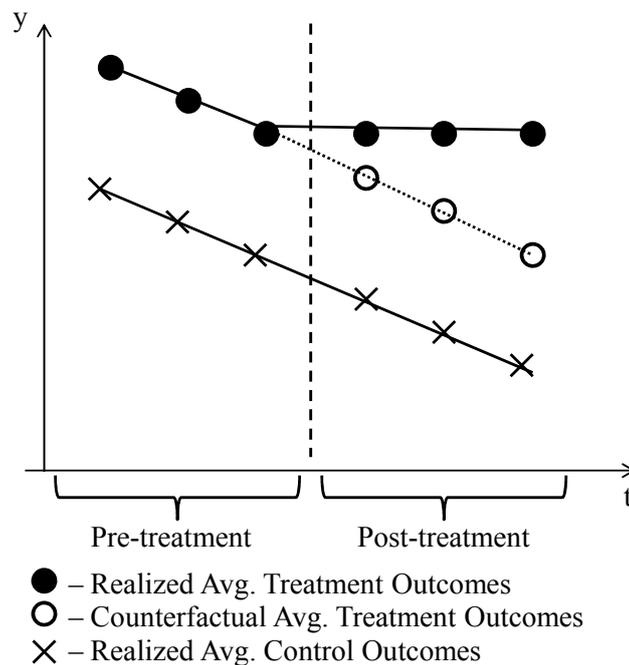
where i , g , and t index subjects, groups, and time periods, respectively, d_{gt} is an indicator identifying whether the treatment has affected group g at time t , X_{igt} a vector of covariates, p_t are period fixed effects, m_g are group fixed effects, and u_{igt} is the error term. The treatment effect is given by β_1 .

4.3.3 The Key Identifying Assumption for DD

The key assumption for consistency of the DD estimator is, as with all regression based estimators, the zero correlation assumption. Economically, this condition means that in the absence of treatment, the average change in the response variable would have been the same for both the treatment and control groups. This assumption is often referred to as the “parallel trends” assumption because it requires any trends in outcomes for the treatment and control groups prior to treatment to be the same.

Figure 1 illustrates this idea by plotting the average treatment and control response functions during the pre- and post-treatment periods. To highlight the concept of parallel trends, we assume that there are three distinct outcomes in the pre- and post-treatment eras. The realized average treatment and control outcomes are represented by the filled circles and “x”s on the solid lines, respectively. The pre- and post-treatment periods are delineated by the vertical dashed line, and indicated on the horizontal axis. The empty circles on the dashed line in the post-treatment period represent the counterfactual outcomes, what would have happened to the subjects in the treatment group had they not received treatment. There are several features of the picture to note.

Figure 1: **Difference-in-Differences Intuition**



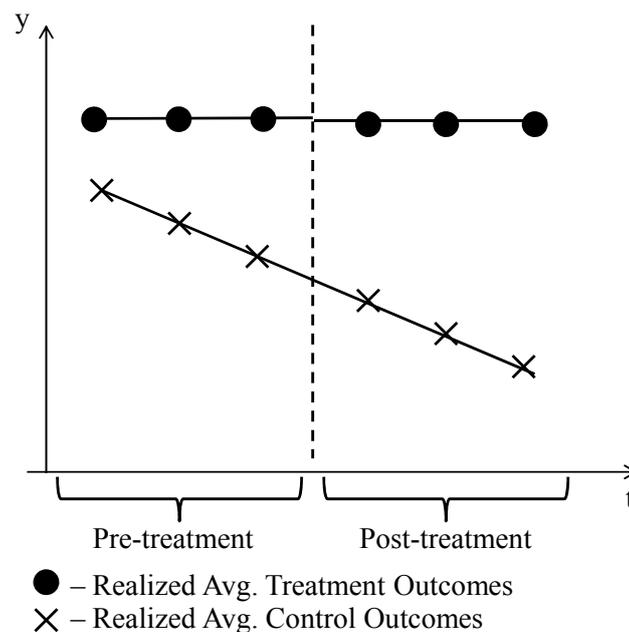
First, the average outcome for the treatment and control groups are different; treated firms have higher outcomes than control firms, on average. Second, outcomes for both treated and control firms appear to be trending down at the same rate during the pre-treatment period. As noted above, neither of these issues are cause for concern since these patterns are fully accounted for by the d and p control variables, respectively.

The third feature to notice is a kink or break in the realized outcome for the treatment group occurring immediately after the onset of the treatment. Rather than continuing on the pre-treatment trend, as indicated by the dashed line representing the counterfactual outcomes, the treatment group appears to have abruptly altered its behavior as a result of

the treatment. This sharp change in behavior among the subjects that are treated—and lack of change in behavior for the control group—is the variation that the DD estimator uses for identifying the treatment effect.

Fourth, the picture highlights the importance of the parallel trends assumption. While level differences and common trends are easily handled by the DD estimator, differential trends among the treatment and control groups will generally lead to inconclusive or erroneous inferences. Figure 2 highlights this point. There is no change in behavior among either treatment or control groups following the onset of treatment. Yet, the DD estimator will estimate a large positive effect simply because of the different trends. This estimated difference could be because the treatment had no effect on the treated subjects, or perhaps the treatment had a significant effect on the treated subjects, whose outcomes may have been significantly lower or higher without intervention. The point is that there is no way of knowing and, therefore, the estimated treatment effect is unidentified.

Figure 2: **Violation of the Parallel Trends Assumption**



As with all endogeneity problems, we cannot formally test the parallel trends assumption, i.e.,

$$\text{cov}(d, u) = \text{cov}(p, u) = \text{cov}(dp, u) = 0.$$

What we can do, assuming we have more than one period's worth of pre-treatment data, is compare the trends in the outcome variables during the pre-treatment era. For example,

Figure 1 clearly passes this visual inspection, while Figure 2 does not. One can even perform a formal statistical test (e.g., paired sample t-test) of the difference in average growth rates across the two groups. However, while similar pre-treatment trends are comforting, indeed necessary, they are not a sufficient condition to ensure that the endogeneity problem has been solved. If there are omitted time-varying variables that differentially affect the treatment and control groups, but are correlated with the outcome variable, then the coefficient estimates will be inconsistent.

Another potential concern arises when the treatment and control groups have different pre-treatment levels for the outcome variable. For example, in Figure 1, the level of the outcome variable for the treatment group is, on average, higher than that for the control group. As noted above, this difference does not compromise the internal validity of the DD estimator because these differences are captured by the d indicator. However, different levels of the outcome variable increases the sensitivity of the DD estimator to the functional form assumption.

To illustrate, consider a case in which the average outcome for the treatment group increases from 40% to 60% between the pre- and post-treatment periods. Likewise, the average outcome for the control group increases from 10% to 20%. The DD estimate is $(60\% - 40\%) - (20\% - 10\%) = 10\%$. However, if we consider the natural log of the outcome variable instead of the level, the DD estimator is: $(\ln(60\%) - \ln(40\%)) - (\ln(20\%) - \ln(10\%)) = -0.29$. This example shows that the outcome increased by more for the treatment group than the control group in absolute terms (20% versus 10%). But, in relative terms, the increase in outcomes for the treatment group is relatively small when compared to the control group (41% versus 70%). Which answer is correct depends upon the question being asked.

4.4 Checking Internal Validity

Because the key assumption behind the DD estimator, the parallel trends assumption, is untestable, a variety of sensitivity or robustness tests should be performed. We present a laundry list of such tests here, though we suspect only a few apply or are necessary for any given study.

- **Falsification Test #1:** Repeat the DD analysis on pre-event years. Falsely assume that the onset of treatment occurs one (two, three, ...) year before it actually does. Repeat the estimation. The estimated treatment effect should be statistically indistinguishable from zero to ensure that the observed change is more likely due to

the treatment, as opposed to some alternative force. This exercise is undertaken by Almeida et al. (2012) who examine the link between corporate debt maturity structure and investment in the wake of the credit crisis. Loosely speaking, one can think of the treatment and control groups in their study as being comprised of firms with a lot or a little debt maturing shortly after the onset of the crisis, respectively. The treatment is the crisis itself, whose onset began in 2007.

A simplified version of their model is

$$y = \beta_0 + \beta_1 d \times p + \beta_2 d + \beta_3 p + u,$$

where y is fixed investment, d equals one for firms in the treatment group and zero for firms in the control group, and p equals one from 2007 onward and zero prior to 2007. By changing the breakpoint defining p from 2007 to 2006 and 2005, they are able to show that the effect is isolated to periods occurring only after the onset of the crisis.

Bertrand and Mullainathan (2003) carry out a similar exercise in their study of wages and state antitakeover laws. A simplified version of their model is

$$y = \beta_0 + \beta_1 d \times p_{-1} + \beta_2 d \times p_0 + \beta_3 d \times p_1 + \beta_4 d + \beta_5 p_{-1} + \beta_6 p_0 + \beta_7 p_1 + u,$$

where y is wages, d equals one if the state in which the firm is registered passed an antitakeover law and zero otherwise, p_1 is an indicator variable equal to one during the period just after the passage of the law, and zero otherwise, p_{-1} is an indicator variable equal to one during the period before the passage of the law, and zero otherwise, p_0 is an indicator variable equal to one during the period in which the law was passed, and zero otherwise, and they show that $\hat{\beta}_1$ is indistinguishable from zero and $\hat{\beta}_2$ is significantly smaller than the estimated treatment effect $\hat{\beta}_3$.

- **Falsification Test #2:** Make sure that variables that should be unaffected by the event are unaffected by the event. To do so, replace the outcome variable of interest in the empirical model, y , (e.g., Eqn (32)) with these other variables.
- **Multiple Treatment and Control Groups:** Multiple treatment and control groups reduce any biases and noise associated with just one comparison. While the treatment and control groups should be similar along outcome relevant dimensions, differences across the groups within each category (treatment or control) are helpful in so far as these differences are likely to come with different biases. Bertrand and Mullainathan (2003) effectively use a variety of treatment and control groups in their analysis since states passed antitakeover laws at different dates. Jayaratne and Strahan (1996) are

similar in their analysis of bank branch deregulations, which occurred at different times for different states.

- **Difference-in-Differences-in-Differences** With multiple treatment or control groups one can perform a triple difference, which is the difference in two difference-in-difference estimates. While useful, the triple difference may be more trouble than it is worth. For example, consider the case with one treatment group and two control groups. If one of the DD estimates is zero, the additional differencing sacrifices statistical power and inflates the standard error on the estimated treatment effect. If the other DD estimate is not zero, then one must wonder about the internal validity of the original DD estimate. Tsoutsoura (2010) and Morse (2011) provide corporate finance examples of triple differences.
- **Balanced Treatment and Control Groups:** The treatment and control groups should be relatively similar along observable dimensions relevant for treatment, i.e., balanced. If not, then incorporating control variables in the regression specification can help. However, if the groups differ significantly along observables, chances are that they may differ along unobservables. Simple pairwise t-tests or a regression of the treatment indicator on observables should reveal no statistically significant results if the groups are balanced.
- **Timing of Behavior Change:** The change in treatment group behavior should be concentrated around the onset of treatment. Moving further away from this event allows other confounding factors to influence outcomes and threaten the internal validity of the study.
- **Control by Systematic Variation:** This notion is related to the use of multiple control groups, and is discussed in Rosenbaum (1987). Imagine that the treatment and control groups differ along some unobservable dimension, call it U . If another control group exists that differs significantly from the first along the unobserved dimension, then this second group can provide a test of the relevance of the omitted variable for the treatment effect. Intuitively, if U is important for the outcome, then variation in U should impact the estimated treatment effect. Because U is unobservable, a direct measure (e.g., explanatory variable) cannot be incorporated into the regression. However, we can use the two control groups as a means to test the relevance of variation in U as a confounder for the treatment effect.
- **Treatment Reversal** If the onset of treatment causes a change in behavior, then, all else equal, the reversal of that treatment should cause a return to the pre-treatment

behavior. For example, Leary (2009) examines an expansion and subsequent contraction in the supply of bank credit using the emergence of the market for certificates of deposit (CDs) in 1961 and the credit crunch of 1966, respectively. He shows that the debt issuances and leverage ratios of bank-dependent borrowers experienced significant increases and decreases relative to non-bank-dependent borrowers in response to the two events.

4.5 Further Reading

Other examples of DD estimations in finance include: Agrawal (2009), Asker and Ljungqvist (2010), Gormley and Matsa (2011), Melzer (2010), and Becker and Stromberg (2010). Textbook treatments of natural experiments and difference-in-differences estimators can be found in Chapters 6 and 10 in Wooldridge (2002), chapters 10 and 13 in Stock and Watson (2007), and chapter 5 of Angrist and Pischke (2010). Meyer (1995) provides an excellent discussion of these topics using the labor literature as an illustrative vehicle, as does the survey of empirical methods in labor economics by Angrist and Krueger (1999). The empirical methods mimeo by Esther Duflo,²⁴ while terse, presents a number of tips and examples of these methods, using examples from labor, development, and public finance.

5. Regression Discontinuity Design

Regression discontinuity design (RDD) is another quasi-experimental technique. Unlike natural experiments, assignment to treatment and control groups in a RDD is *not* random. Instead, RDD takes advantage of a known cutoff or threshold determining treatment assignment or the probability of receiving treatment. Assuming these assignment rules are functions of one or more observable continuous variables, the cutoff generates a discontinuity in the treatment recipiency rate at that point. Recipients whose assignment variable is above the cutoff are assigned to one group (e.g., treatment), those below assigned to the other (e.g., control). To be clear, in this context “recipients” refers to any economic agents—firms, managers, investors—that are affected by the assignment rule.

For example, Chava and Roberts (2008) examine the link between financing and investment using violations of financial covenants in bank loans. Financial covenants specify thresholds for certain accounting variables above or below which a firm is deemed to be in vi-

²⁴See <http://piketty.pse.ens.fr/fichiers/enseig/econeg/article/Duflo2002.pdf>.

olation of the contract terms. These covenant thresholds provide a deterministic assignment rule distinguishing treatment (in violation) and control (not in violation) groups.²⁵

Another feature of RDD that distinguishes it from natural experiments is that one need not assume that the cutoff generates variation that is as good as randomized. Instead, randomized variation is a consequence of RDD so long as agents are unable to precisely control the assignment variable(s) near the cutoff (Lee, 2008). For example, if firms could perfectly manipulate the net worth that they report to lenders, or if consumers could perfectly manipulate their FICO scores, then a RDD would be inappropriate in these settings. More broadly, this feature makes RDD studies particularly appealing because they rely on relatively mild assumptions compared to other non-experimental techniques (Lee and Lemieux, 2010).

There are several other appealing features of RDD. RDDs abound once one looks for them. Program resources are often allocated based on formulas with cutoff structures. RDD is intuitive and often easily conveyed in a picture, much like the difference-in-differences approach. In RDD, the picture shows a sharp change in outcomes around the cutoff value, much like the difference-in-differences picture shows a sharp change in outcomes for the treatment group after the event.

The remainder of this section will outline the RDD technique, which comes in two flavors: sharp and fuzzy. We first clarify the distinction between the two. We then discuss how to implement RDD in practice. Unfortunately, applications of RDD in corporate finance are relatively rare. Given the appeal of RDD, we anticipate that this dearth will change in the coming years. For now, we focus attention on the few existing studies, occasionally referring to examples from the labor literature as needed, in order to illustrate the concepts discussed below.

5.1 Sharp RDD

In a sharp RDD, subjects are assigned to or selected for treatment solely on the basis of a cutoff value of an observed variable.²⁶ This variable is referred to by a number of names in

²⁵Other corporate finance studies incorporating RDDs include Keys et al. (2010), which examines the link between securitization and lending standards using a guideline established by Fannie Mae and Freddie Mac that limits securitization to loans to borrowers with FICO scores above a certain limit. This rule generates a discontinuity in the probability of securitization occurring precisely at the 620 FICO score threshold. In addition, Baker and Xuan (2009) examine the role reference points play in corporate behavior, Roberts and Sufi (2009a) examine the role of covenant violations in shaping corporate financial policy, and Black and Kim (2011) examine the effects on corporate governance of a rule stipulating a minimum fraction of outside directors.

²⁶The requirement that the variable be observable rules out situations, such as accounting disclosure rules, in which the variable is observable on only one side of the cutoff.

the econometrics literature including: assignment, forcing, selection, running, and ratings. In this paper, we will use the term forcing. The forcing variable can be a single variable, such as a borrower’s FICO credit score in Keys et al. (2010) or a firm’s net worth as in Chava and Roberts (2008). Alternatively, the forcing variable can be a function of a single variable or several variables.

What makes a sharp RDD sharp is the first key assumption.

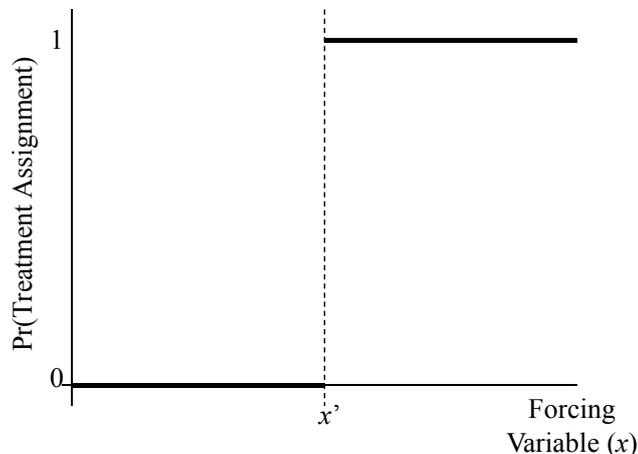
Sharp RDD Key Assumption # 1: Assignment to treatment occurs through a known and measured *deterministic* decision rule:

$$d = d(x) = \begin{cases} 1 & \text{if } x \geq x' \\ 0 & \text{otherwise.} \end{cases} \quad (34)$$

where x is the forcing variable and x' the threshold.

In other words, assignment to treatment occurs if the value of the forcing variable x meets or exceeds the threshold x' .²⁷ Graphically, the assignment relation defining a sharp RDD is displayed in Figure 3, which has been adapted from Figure 3 in Imbens and Lemieux (2008). In the context of Chava and Roberts (2008), when a firm’s debt-to-EBITDA ratio,

Figure 3: **Probability of Treatment Assignment in Sharp RDD**



for example, (x) rises above the covenant threshold (x') , the firm’s state changes from not in violation (control) to in violation (treatment) with certainty.

²⁷We refer to a scalar variable x and threshold x' only to ease the discussion. The weak inequality is unimportant since x is assumed to be continuous and therefore $Pr(x = x') = 0$. The direction of the inequality is unimportant, arbitrarily chosen for illustrative purposes. However, we do assume that x has a positive density in the neighborhood of the cutoff x' .

5.1.1 Identifying Treatment Effects

Given the delineation of the data into treatment and control groups by the assignment rule, a simple, albeit naive, approach to estimation would be a comparison of sample averages. As before, this comparison can be accomplished with a simple regression

$$y = \alpha + \beta d + u \tag{35}$$

where $d = 1$ for treatment observations and zero otherwise. However, this specification assumes that treatment assignment d and the error term u are uncorrelated so that assignment is as if it is random with respect to potential outcomes.

In the case of RDD, assignment is determined by a known rule that ensures treatment assignment is correlated with the forcing variable, x , so that d is almost surely correlated with u and OLS will not recover a treatment effect of interest (e.g., ATE, ATT). For example, firms' net worths and current ratios (i.e., current assets divided by current liabilities), are the forcing variables in Chava and Roberts (2008). A comparison of investment between firms in violation of their covenants and those not in violation will, by construction, be a comparison of investment between two groups of firms with very different net worths and current ratios. However, the inability to precisely measure marginal q may generate a role for these accounting measures in explaining fixed investment (Erickson and Whited, 2000; Gomes, 2001). In other words, the comparison of sample averages is confounded by the forcing variables, net worth and current ratio.²⁸

One way to control for x is to include it in the regression as another covariate:

$$y = \alpha + \beta d + \gamma x + u \tag{36}$$

However, this approach is also unappealing because identification of the parameters comes from all of the data, including those points that are far from the discontinuity. Yet, the variation on which RDD relies for proper identification of the parameters is that occurring precisely at the discontinuity. This notion is formalized in the second key assumption of sharp RDD, referred to as the **local continuity** assumption.

²⁸One might think that matching would be appropriate in this instance since a sharp RDD is just a special case of selection on observables (Heckman and Robb, 1985). However, in this setting there is a violation of the second strong ignorability conditions (Rosenbaum and Rubin, 1983), which require (1) that u be independent of d conditional on x (unconfoundedness), and (2) that $0 < Pr(d = 1|x) < 1$ (overlap). Clearly, the overlap assumption is a violation since $Pr(d = 1|x) \in \{0, 1\}$. In other words, at each x , every observation is either in the treatment or control group, but never both.

RDD Key Assumption # 2: Both potential outcomes, $E(y(0)|x)$ and $E(y(1)|x)$, are continuous in x at x' . Equivalently, $E(u|x)$ is continuous in x at x' .²⁹

Local continuity is a general assumption invoked in both sharp and fuzzy RDD. As such, we do not preface this assumption with “Sharp,” as in the previous assumption.

Assuming a positive density of x in a neighborhood containing the threshold x' , local continuity implies that the limits of the conditional expectation function around the threshold recover the ATE at x' . Taking the difference between the left and right limits in x of Eqn (35) yields,

$$\begin{aligned} \lim_{x \downarrow x'} E(y|x) - \lim_{x \uparrow x'} E(y|x) &= \left[\lim_{x \downarrow x'} E(\beta d|x) + \lim_{x \downarrow x'} E(u|x) \right] - \left[\lim_{x \uparrow x'} E(\beta d|x) + \lim_{x \uparrow x'} E(u|x) \right] \\ &= \beta, \end{aligned} \tag{37}$$

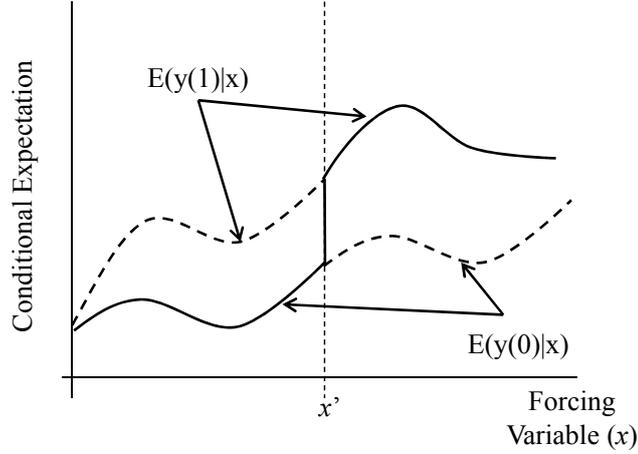
where the second line follows because continuity implies that $\lim_{x \downarrow x'} E(u|x) - \lim_{x \uparrow x'} E(u|x) = 0$. In other words, a comparison of average outcomes just above and just below the threshold identifies the ATE for subjects sufficiently close to the threshold. Identification is achieved assuming only smoothness in expected potential outcomes at the discontinuity. There are no parametric functional form restrictions.

Consider Figure 4, which is motivated from Figure 2 of Imbens and Lemieux (2008). On the vertical axis is the conditional expectation of outcomes, on the horizontal axis the forcing variable. Conditional expectations of potential outcomes, $E(y(0)|x)$ and $E(y(1)|x)$, are represented by the continuous curves, part of which are solid and part of which are dashed. The solid parts of the curve correspond to the regions in which the potential outcome is observed, and the dashed parts are the counterfactual. For example, $y(1)$ is observed only when the forcing variable is greater than the threshold and the subject is assigned to treatment. Hence, the part of the curve to the right of x' is solid for $E(y(1)|x)$ and dashed for $E(y(0)|x)$.

The local continuity assumption is that the conditional expectations representing potential outcomes are smooth (i.e., continuous) around the threshold, as illustrated by the figure. What this continuity ensures is that the average outcome is similar for subjects close to but on different sides of the threshold. In other words, in the absence of treatment, outcomes would be similar. However, the conditional expectation of the observed outcome, $E(y|x)$, is

²⁹This is a relatively weak but unrealistic assumption as continuity is only imposed at the threshold. As such, two alternative, stronger assumptions are sometimes made. The first is continuity of conditional regression functions, such that $E(y(0)|x)$ and $E(y(1)|x)$ are continuous in $x, \forall x$. The second is continuity of conditional distribution functions, such that $F(y(0)|x)$ and $F(y(1)|x)$ are continuous in $x, \forall x$.

Figure 4: **Conditional Expectation of Outcomes in Sharp RDD**



represented by the all solid line that is discontinuous at the threshold, x' . Thus, continuity ensures that the only reason for different outcomes around the threshold is the treatment.

While a weak assumption, local continuity does impose limitations on inference. For example, consider a model with heterogeneous effects,

$$y = \alpha + \tilde{\beta}d + u \quad (38)$$

where $\tilde{\beta}$ is a random variable that can vary with each subject. In this case, we also require local continuity of $E(\tilde{\beta}|x)$ at x' . Though we can identify the treatment effect under this assumption, we can only learn about that effect for the subpopulation that is close to the cutoff. This may be a relatively small group, suggesting little external validity. Further, internal validity may be threatened if there are coincidental functional discontinuities. One must be sure that there are no other confounding forces that induce a discontinuity in the outcome variable coincident with that induced by the forcing variable of interest.

5.2 Fuzzy RDD

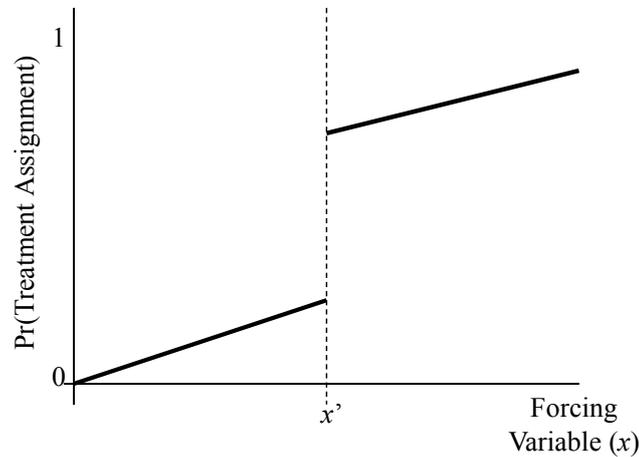
The primary distinction from a sharp RDD is captured by the first key assumption of a fuzzy RDD.

Fuzzy RDD Key Assumption # 1: Assignment to treatment occurs in a *stochastic* manner where the probability of assignment (a.k.a. **propensity score**) has a known discontinuity at x' .

$$0 < \lim_{x \downarrow x'} Pr(d = 1|x) - \lim_{x \uparrow x'} Pr(d = 1|x) < 1.$$

Instead of a 0-1 step function, as in the sharp RDD case, the treatment probability as a function of x in a fuzzy RDD can contain a jump at the cutoff that is less than one. This situation is illustrated in Figure 5, which is analogous to figure 3 in the sharp RDD case.

Figure 5: **Probability of Treatment Assignment in Fuzzy RDD**



An example of a fuzzy RDD is given in Keys et al. (2010). Loans with FICO scores above 620 are only more likely to be securitized. Indeed securitization occurs both above and below this threshold. Thus, one can also think of fuzzy RDD as akin to mis-assignment relative to the cutoff value in a sharp RDD. This mis-assignment could be due to the use of additional variables in the assignment that are unobservable to the econometrician. In this case, values of the forcing variable near the cut-off appear in both treatment and control groups.³⁰ Likewise, Bakke, et al. (in press) is another example of a fuzzy RDD because some of the causes of delisting, such as governance violations, are not observable to the econometrician. Practically speaking, one can imagine that the incentives to participate in the treatment change discontinuously at the cutoff, but they are not powerful enough to move all subjects from non-participant to participant status.

In a fuzzy RDD one would not want to compare the average outcomes of treatment and control groups, even those close to the threshold. The fuzzy aspect of the RDD suggests that subjects may self-select around the threshold and therefore be very different with respect to unobservables that are relevant for outcomes. To illustrate, reconsider the Bakke, et al. (in press) study. First, comparing firms that delist to those that do not delist is potentially

³⁰Fuzzy RDD is also akin to random experiments in which there are members of the treatment group that do not receive treatment (i.e., “no-shows”), or members of the control group who do receive treatment (i.e., “cross-overs”).

confounded by unobserved governance differences, which are likely correlated with outcomes of interest (e.g., investment, financing, employment, etc.).

5.2.1 Identifying Treatment Effects

Maintaining the assumption of local continuity and a common treatment effect,

$$\begin{aligned} \lim_{x \downarrow x'} E(y|x) - \lim_{x \uparrow x'} E(y|x) &= \left[\lim_{x \downarrow x'} E(\beta d|x) + \lim_{x \downarrow x'} E(u|x) \right] - \left[\lim_{x \uparrow x'} E(\beta d|x) + \lim_{x \uparrow x'} E(u|x) \right] \\ &= \beta \left[\lim_{x \downarrow x'} E(d|x) - \lim_{x \uparrow x'} E(d|x) \right]. \end{aligned}$$

This result implies that the treatment effect, common to the population, β , is identified by

$$\beta = \frac{\lim_{x \downarrow x'} E(y|x) - \lim_{x \uparrow x'} E(y|x)}{\lim_{x \downarrow x'} E(d|x) - \lim_{x \uparrow x'} E(d|x)}. \quad (39)$$

In other words, the common treatment effect is a ratio of differences. The numerator is the difference in expected outcomes near the threshold. The denominator is the change in probability of treatment near the threshold. The denominator is always non-zero because of the assumed discontinuity in the propensity score function (Fuzzy RDD Key Assumption #1). Note that Eqn (39) is equal to Eqn (37) when the denominator equals one. This condition is precisely the case in a sharp RDD. (See Sharp RDD Key Assumption #1.)

When the treatment effect is not constant, $\tilde{\beta}$, we must maintain that $E(\tilde{\beta}|x)$ is locally continuous at the threshold, as before. In addition, we must assume local conditional independence of $\tilde{\beta}$ and d , which requires d to be independent of $\tilde{\beta}$ conditional on x near x' (Hahn, Todd, and van der Klaauw, 2001). In this case,

$$\begin{aligned} \lim_{x \downarrow x'} E(y|x) - \lim_{x \uparrow x'} E(y|x) &= \left[\lim_{x \downarrow x'} E(\tilde{\beta} d|x) + \lim_{x \downarrow x'} E(u|x) \right] - \left[\lim_{x \uparrow x'} E(\tilde{\beta} d|x) + \lim_{x \uparrow x'} E(u|x) \right] \\ &= \lim_{x \downarrow x'} E(\tilde{\beta}|x) \lim_{x \downarrow x'} E(d|x) - \lim_{x \uparrow x'} E(\tilde{\beta}|x) \lim_{x \uparrow x'} E(d|x) \end{aligned}$$

By continuity of $E(\tilde{\beta}|x)$, this result implies that the ATE can be recovered with the same ratio as in Eqn (39). That is,

$$E(\tilde{\beta}|x) = \frac{\lim_{x \downarrow x'} E(y|x) - \lim_{x \uparrow x'} E(y|x)}{\lim_{x \downarrow x'} E(d|x) - \lim_{x \uparrow x'} E(d|x)}. \quad (40)$$

The practical problem with heterogeneous treatment effects involves violation of the conditional independence assumption. If subjects self-select into treatment or are selected on

the basis of expected gains from the treatment, then this assumption is clearly violated. That is, the treatment effect for individuals, $\tilde{\beta}$ is not independent of the treatment assignment, d . In this case, we must employ an alternative set of assumptions to identify an alternative treatment effect called a local average treatment effect (LATE) (Angrist and Imbens, 1994).

Maintaining the assumptions of (1) discontinuity in the probability of treatment, (2) local continuity in potential outcomes, identification of LATE requires two additional assumptions (Hahn, Todd, and van der Klaauw, 2001). First, $(\beta, d(x))$ is jointly independent of x near x' , where $d(x)$ is a deterministic assignment rule that varies across subjects. Second,

$$\exists \epsilon > 0 : d(x' + \delta) \geq d(x' - \delta) \forall 0 < \delta < \epsilon.$$

Loosely speaking, this second condition requires that the likelihood of treatment assignment always be weakly greater above the threshold than below. Under these conditions, the now familiar ratio,

$$\frac{\lim_{x \downarrow x'} E(y|x) - \lim_{x \uparrow x'} E(y|x)}{\lim_{x \downarrow x'} E(d|x) - \lim_{x \uparrow x'} E(d|x)}, \quad (41)$$

identifies the LATE, which is defined as

$$\lim_{\delta \rightarrow 0} E(\beta | d(x' + \delta) - d(x' - \delta) = 1). \quad (42)$$

The LATE represents the average treatment effect of the compliers, that is, those subjects whose treatment status would switch from non-recipient to recipient if their score x crossed the cutoff. The share of this group in the population in the neighborhood of the cutoff is just the denominator in Eqn (41).

Returning to the delisting example from Bakke, et al. (in press), assume that delisting is based on the firm's stock price relative to a cutoff and governance violations. In other words, all firms with certain governance violations are delisted and only those non-violating firms with sufficiently low stock prices are delisted. If governance violations are unobservable, then the delisting assignment rule generates a fuzzy RDD, as discussed above. The LATE applies to the subgroup of firms with stock prices close to the cutoff for whom delisting depends on their stock price's position relative to the cutoff, i.e., non-violating firms. For more details on these issues, see studies by van der Klaauw (2008) and Chen and van der Klaauw (2008) that examine the economics of education and scholarship receipt.

5.3 Graphical Analysis

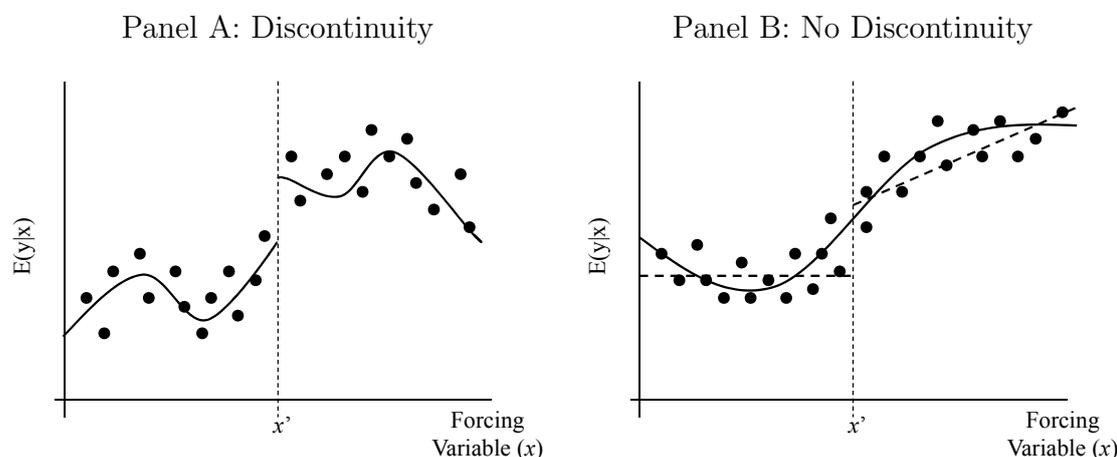
Perhaps the first place to start in analyzing a RDD is with some pictures. For example, a plot of $E(y|x)$ is useful to identify the presence of a discontinuity. To approximate this conditional expectation, divide the domain of x into bins, as one might do in constructing a histogram. Care should be taken to ensure that the bins fall on either side of the cutoff x' , and no bin contains x' in its interior. Doing so ensures that treatment and control observations are not mixed together into one bin by the researcher, though this may occur naturally in a fuzzy RDD. For each bin, compute the average value of the outcome variable y and plot this value above the mid-point of the bin.

Figure 6 presents two hypothetical examples using simulated data to illustrate what to look for (Panel A) and what to look out for (Panel B).³¹ Each circle in the plots corresponds to the average outcome, y , for a particular bin that contains a small range of x -values. We also plot estimated regression lines in each panel. Specifically, we estimate the following regression in Panel A,

$$y = \alpha + \beta d + \sum_{s=1}^5 [\beta_s x^s + \gamma_s d \cdot x^s] + u$$

and a cubic version in Panel B.

Figure 6: RDD Examples



Focusing first on Panel A, there are several features to note, as suggested by Lee and Lemieux (2010). First, the graph provides a simple means of visualizing the functional

³¹Motivation for these figures and their analysis comes from Chapter 6 of Angrist and Pischke (2009).

form of the regression, $E(y|x)$ because the bin means are the nonparametric estimate of the regression function. In Panel A, we note that a fifth-order polynomial is needed to capture the features of the conditional expectation function. Further, the fitted line reveals a clear discontinuity. In contrast, in Panel B a cubic, maybe a quadratic, polynomial is sufficient and no discontinuity is apparent.

Second, a sense of the magnitude of the discontinuity can be gleaned by comparing the mean outcomes in the two bins on both sides of the threshold. In Panel A, this magnitude is represented by the jump in $E(y|x)$ moving from just below x' to just above. Panel B highlights the importance of choosing a flexible functional form for the conditional expectation. Assuming a linear functional form, as indicated by the dashed lines, would incorrectly reveal a discontinuity. In fact, the data reveal a nonlinear relation between the outcome and forcing variables.

Finally, the graph can also show whether there are similar discontinuities in $E(y|x)$ at points other than x' . At a minimum, the existence of other discontinuities requires an explanation to ensure that what occurs at the threshold is in fact due to the treatment and not just another “naturally occurring” discontinuity.

As a practical matter, there is a question of how wide the bins should be. As with most nonparametrics, this decision represents a tradeoff between bias and variance. Wider bins will lead to more precise estimates of $E(y|x)$, but at the cost of bias since wide bins fail to take into account the slope of the regression line. Narrower bins mitigate this bias, but lead to noisier estimates as narrow bins rely on less data. Ultimately, the choice of bin width is subjective but should be guided by the goal of creating a figure that aids in the analysis used to estimate treatment effects.

Lee and Lemieux (2010) suggest two approaches. The first is based on a standard regression F-test. Begin with some number of bins denoted K and construct indicator variables identifying each bin. Then divide each bin in half and construct another set of indicator variables denoting these smaller bins. Regress y on the smaller bin indicator variables and conduct an F-test to see if the additional regressors (i.e., smaller bins) provide significant additional explanatory power. If not, then the original K bins should be sufficient to avoid oversmoothing the data.

The second test adds a set of interactions between the bin dummies, discussed above, and the forcing variable, x . If the bins are small enough, there should not be a significant slope within each bin. Recall that plotting the mean outcome above the midpoint of each bin presumes an approximately zero slope within the bin. A simple test of this hypothesis is a joint F-test of the interaction terms.

In the case of fuzzy RDD, it can also be useful to create a similar graph for the treatment dummy, d_i , instead of the outcome variable. This graph can provide an informal way of estimating the magnitude of the discontinuity in the propensity score at the threshold. The graph can also aid with the choice of functional form for $E(d|x) = Pr(d|x)$.

Before discussing estimation, we mention a caveat from Lee and Lemieux (2010). Graphical analysis can be helpful but should not be relied upon. There is too much room for researchers to construct graphs in a manner that either conveys the presence of treatment effects when there are none, or masks the presence of treatment effects when they exist. Therefore, graphical analysis should be viewed as a tool to guide the formal estimation, rather than as a necessary or sufficient condition for the existence of a treatment effect.

5.4 Estimation

As is clear from Eqns (37), (40), and (41), estimation of various treatment effects requires estimating boundary points of conditional expectation functions. Specifically, we need to estimate four quantities:

- 1) $\lim_{x \downarrow x'} E(y_i|x)$,
- 2) $\lim_{x \uparrow x'} E(y_i|x)$,
- 3) $\lim_{x \downarrow x'} E(d_i|x)$, and
- 4) $\lim_{x \uparrow x'} E(d_i|x)$.

The last two quantities are only relevant for the fuzzy RDD, since a sharp design assumes that $\lim_{x \downarrow x'} E(d_i|x) = 1$ and $\lim_{x \uparrow x'} E(d_i|x) = 0$.

5.4.1 Sharp RDD

In theory, with enough data one could focus on the area just around the threshold, and compare average outcomes for these two groups of subjects. In practice, this approach is problematic because a sufficiently small region will likely run into power problems. As such, widening the area of analysis around the threshold to mitigate power concerns is often necessary. Offsetting this benefit of extrapolation is an introduction of bias into the estimated treatment effect as observations further from the discontinuity are incorporated into the estimation. Thus, the tradeoff researchers face when implementing a RDD is a common one: bias versus variance.

One way of approaching this problem is to emphasize power by using all of the data and to try to mitigate the bias through observable control variables, and in particular the forcing variable, x . For example, one could estimate two separate regressions on each side of the cutoff point:

$$y_i = \beta^b + f(x_i - x') + \varepsilon_i^b \quad (43)$$

$$y_i = \beta^a + g(x_i - x') + \varepsilon_i^a \quad (44)$$

where the superscripts denote below (“b”) and above (“a”) the threshold, x' , and f and g are continuous functions (e.g., polynomials). Subtracting the threshold from the forcing variable means that the estimated intercepts will provide the value of the regression functions at the threshold point, as opposed to zero. The estimate treatment effect is just the difference between the two estimated intercepts, $(\beta_a - \beta_b)$.³²

An easier way to perform inference is to combine the data on both sides of the threshold and estimate the following pooled regression:

$$y_i = \alpha + \beta d_i + f(x_i - x') + d_i \cdot g(x_i - x') + \varepsilon_i \quad (45)$$

where f and g are continuous functions. The treatment effect is β , which equals $(\beta^a - \beta^b)$. Note, this approach maintains the functional form flexibility associated with estimating two separate regressions by including the interaction term $d_i \cdot g(x_i - x')$. This is an important feature since there is rarely a strong a priori rationale for constraining the functional form to be the same on both sides of the threshold.³³

The functions f and g can be specified in a number of ways. A common choice is polynomials. For example, if f and g are quadratic polynomials, then Eqn (45) is:

$$y_i = \alpha + \beta d_i + \gamma_1^b(x_i - x') + \gamma_2^b(x_i - x')^2 + \gamma_3^a d_i(x_i - x') + \gamma_4^a d_i(x_i - x')^2 + \varepsilon_i$$

This specification fits a different quadratic curve to observations above and below the threshold. The regression curves in Figure 6 are an example of this approach using quintic (and cubic) polynomials for f and g .

An important consideration when using a polynomial specification is the choice of polynomial order. While some guidance can be obtained from the graphical analysis, the correct

³²This approach of incorporating controls for x as a means to correct for selection bias due to selection on observables is referred to as the **control function** approach (Heckman and Robb, 1985). A drawback of this approach is the reduced precision in the treatment effect estimate caused by the collinearity between d_i and f and g . This collinearity reduces the independent variation in the treatment status and, consequently, the precision of the treatment effect estimates (van der Klaauw, 2008).

³³There is a benefit of increased efficiency if the restriction is correct. Practically speaking, the potential bias associated with an incorrect restriction likely outweighs any efficiency gains.

order is ultimately unknown. There is some help from the statistics literature in the form of generalized cross-validation procedures (e.g., van der Klaauw, 2002; Black, Galdo, and Smith, 2007), and the joint test of bin indicators described in Lee and Lemieux (2010). This ambiguity suggests the need for some experimentation with different polynomial orders to illustrate the robustness of the results.

An alternative to the polynomial approach is the use of local linear regressions. Hahn, Todd, and van der Klaauw (2001) show that they provide a nonparametric way of consistently estimating the treatment effect in an RDD. Imbens and Lemieux (2008) suggest estimating linear specifications on both sides of the threshold while restricting the observations to those falling within a certain distance of the threshold (i.e., bin width).³⁴ Mathematically, the regression model is

$$y_i = \alpha + \beta d_i + \gamma_1^b(x_i - x') + \gamma_2^a d_i(x_i - x') + \varepsilon_i, \text{ where } x' - h \leq x \leq x' + h \quad (46)$$

for $h > 0$. The treatment effect is β .

As with the choice of polynomial order, the choice of window width (bandwidth), h , is subjective. Too wide a window increases the accuracy of the estimate, by including more observations, but at the risk of introducing bias. Too narrow a window and the reverse occurs. Fan and Gijbels, 1996) provide a rule of thumb method for estimating the optimal window width. Ludwig and Miller (2007) and Imbens and Lemieux (2008) propose alternatives based on cross-validation procedures. However, much like the choice of polynomial order, it is best to experiment with a variety of window widths to illustrate the robustness of the results.

Of course, one can combine both polynomial and local regression approaches by searching for the optimal polynomial for each choice of bandwidth. In other words, one can estimate the following regression model

$$y_i = \alpha + \beta d_i + f(x_i - x') + d_i \cdot g(x_i - x') + \varepsilon_i, \text{ where } x' - h \leq x \leq x' + h \quad (47)$$

for several choices of $h > 0$, choosing the optimal polynomial order for each choice of h based on one of the approaches mentioned earlier.

One important intuitive point that applies to all of these alternative estimation methods is the tradeoff between bias and efficiency. For example, in terms of the Chava and Roberts (2008) example, Eqn (36) literally implies that the only two variables that are relevant for investment are a bond covenant violation and the distance to the cutoff point for a bond

³⁴This local linear regression is equivalent to a nonparametric estimation with a rectangular kernel. Alternative kernel choices may improve efficiency, but at the cost of less transparent estimation approaches. Additionally, the choice of kernel typically has little impact in practice.

covenant violation. Such a conclusion is, of course, extreme, but it implies that the error term in this regression contains many observable and unobservable variables. Loosely speaking, as long as none of these variables are discontinuous at the exact point of a covenant violation, estimating the treatment effect on a small region around the cutoff does not induce bias. In this small region RDD has both little bias and low efficiency. On the other hand, this argument no longer follows when one uses a large sample, so it is important to control for the differences in characteristics between those observations that are near and far from the cutoff. In this case, because it is nearly impossible in most corporate finance applications to include all relevant characteristics, using RDD on a large sample can result in both high efficiency but also possibly large bias.

One interesting result from Chava and Roberts (2008) that mitigates this second concern is that the covenant indicator variable is largely orthogonal to the usual measures of investment opportunities. Therefore, even though it is hard to control for differences between firms near and far from the cutoff, this omitted variables problem is unlikely to bias the coefficient on the covenant violation indicator. In contrast, in Bakke, et al. (in press) the treatment indicator is not orthogonal to the usual measures of investment opportunities; so inference can only be drawn for the sample of firms near the cutoff and cannot be extrapolated to the rest of the sample. In general, checking orthogonality of the treatment indicator to other important regression variables is a useful diagnostic.

5.4.2 Fuzzy RDD

In a fuzzy RDD, the above estimation approaches are typically inappropriate. When the fuzzy RDD arises because of misassignment relative to the cutoff, $f(x - x')$ and $g(x - x')$ are inadequate controls for selection biases.³⁵ More generally, the estimation approaches discussed above will not recover unbiased estimates of the treatment effect because of correlation between the assignment variable d_i and ε . Fortunately, there is an easy solution to this problem based on instrumental variables.

Recall that including f and g in Eqn (45) helps mitigate the selection bias problem. We can take a similar approach here in solving the selection bias in the assignment indicator, d_i , using the discontinuity as an instrument. Specifically, the probability of treatment can be written as,

$$E(d_i|x_i) = \delta + \phi T + g(x - x') \tag{48}$$

³⁵An exception is when the assignment error is random, or independent of ε conditional on x (Cain, 1975).

where T is an indicator equal to one if $x \geq x'$ and zero otherwise, and g a continuous function. Note that the indicator T is not equal to d_i in the fuzzy RDD because of misassignment or unobservables. Rather,

$$d_i = Pr(d_i|x_i) + \omega$$

where ω is a random error independent of x . Therefore, a fuzzy RDD can be described by a two equation system:

$$y_i = \alpha + \beta d_i + f(x_i - x') + \varepsilon_i, \quad (49)$$

$$d_i = \delta + \phi T_i + g(x_i - x') + \omega_i. \quad (50)$$

Estimation of this system can be carried out with two stage least squares, where d_i is the endogenous variable in the outcome equation and T_i is the instrument. The standard exclusion restriction argument applies: T_i is only relevant for outcomes, y_i , through its impact on assignment, d_i . The estimated β will be equal to the average local treatment effect, $E(\beta_i|x')$. Or, if one replaces the local independence assumption with the local monotonicity condition of Angrist and Imbens, (1994), β estimates the LATE.

The linear probability model in Eqn (50) may appear restrictive, but g (and f) are unrestricted on both sides of the discontinuity, permitting arbitrary nonlinearities. However, one must now choose two bandwidths and polynomial orders corresponding to each equation. Several suggestions for these choices have arisen (e.g., Imbens and Lemieux, 2008). However, practical considerations suggest choosing the same bandwidth and polynomial order for both equations. This restriction eases the computation of the standard errors, which can be obtained from most canned 2SLS routines. It also cuts down on the number of parameters to investigate since exploring different bandwidths and polynomial orders to illustrate the robustness of the results is recommended.

5.4.3 Semiparametric Alternatives

We focused on parametric estimation above by specifying the control functions f and g as polynomials. The choice of polynomial order, or bandwidth, is subjective. As such, we believe that robustness to these choices can be fairly compelling. However, for completeness, we briefly discuss several alternative nonparametric approaches to estimating f and g here. Interested readers are referred to the original articles for further details.

Van der Klaauw (2002) uses a power series approximation for estimating these functions, where the number of power functions is estimated from the data by generalized cross-validation as in Newey et al., (1990). Hahn, Todd, and van der Klaauw (2001) consider

kernel methods using Nadaraya-Watson estimators to estimate the right- and left-hand side limits of the conditional expectations in Eqn (39). While consistent and more robust than parametric estimators, kernel estimators suffer from poor asymptotic bias behavior when estimating boundary points.³⁶ This drawback is common to many nonparametric estimators. Alternatives to kernel estimators that improve upon boundary value estimation are explored by Hahn, Todd, and van der Klaauw (2001) and Porter (2003), both of whom suggest using local polynomial regression (Fan, 1992; Fan and Gijbels, 1996).

5.5 Checking Internal Validity

We have already mentioned some of the most important checks on internal validity, namely, showing the robustness of results to various functional form specifications and bandwidth choices. This section lists a number of additional checks. As with the checks for natural experiments, we are not advocating that every study employing a RDD perform all of the following tests. Rather, this list merely provides a menu of options.

5.5.1 Manipulation

Perhaps the most important assumption behind RDD is local continuity. In other words, the potential outcomes for subjects just below the threshold is similar to those just above the threshold (e.g., see Figure 3). As such, an important consideration is the ability of subjects to manipulate the forcing variable and, consequently, their assignment to treatment and control groups. If subjects can manipulate their value of the forcing variable or if administrators (i.e., those who assign subjects to treatment) can choose the forcing variable or its threshold, then local continuity may be violated. Alternatively, subjects on different sides of the threshold, no matter how close, may not be comparable because of sorting.

For this reason, it is crucial to examine and discuss agents' and administrators' incentives and abilities to affect the values of the forcing variable. However, as Lee and Lemieux (2010) note, manipulation of the forcing variable is not de facto evidence invalidating an RDD. What is crucial is that agents cannot *precisely* manipulate the forcing variable. Chava and Roberts (2008) provide a good example to illustrate these issues.

³⁶As van der Klaauw (2008) notes, if f has a positive slope near x' , the average outcome for observations just to the right of the threshold will typically provide an upward biased estimate of $\lim_{x \downarrow x'} E(y_i|x)$. Likewise, the average outcome of observations just to the left of the threshold would provide a downward biased estimate of $\lim_{x \uparrow x'} E(y_i|x)$. In a sharp RDD, these results generate a positive finite sample bias.

Covenant violations are based on financial figures reported by the company, which has a clear incentive to avoid violating a covenant if doing so is costly. Further, the threshold is chosen in a bargaining process between the borrower and the lender. Thus, possible manipulation is present in both regards: both agents (borrowers) and administrators (lenders) influence the forcing variable and threshold.

To address these concerns, Chava and Roberts (2008) rely on institutional details and several tests. First, covenant violations are not determined from SEC filings, but from private compliance reports submitted to the lender. These reports often differ substantially from publicly available numbers and frequently deviate from GAAP conventions. These facts mitigate the incentives of borrowers to manipulate their reports, which are often shielded from public view because of the inclusion of material nonpublic information. Further mitigating the ability of borrowers to manipulate their compliance reports is the repeated nature of corporate lending, the importance of lending relationships, and the expertise and monitoring role of relationship lenders. Thus, borrowers cannot precisely manipulate the forcing variable, nor is it in their interest to do so.

Regarding the choice of threshold by the lender and borrower, the authors show that violations occur on average almost two years after the origination of the contract. So, this choice would have to contain information about investment opportunities two years hence, which is not contained in more recent measures. While unlikely, the authors include the covenant threshold as an additional control variable, with no effect on their results.

Finally, the authors note that any manipulation is most likely to occur when investment opportunities are very good. This censoring implies that observed violations tend to occur when investment opportunities are particularly poor, so that the impact on investment of the violation is likely understated (see also Roberts and Sufi (2009a)). Further, the authors show that when firms violate, they are more likely to violate by a small amount than a large amount. This is at odds with the alternative that borrowers manipulate compliance reports by “plugging the dam” until conditions get so bad that violation is unavoidable.

A more formal two-step test is suggested by McCrary (2008). The first step of this procedure partitions the forcing variable into equally-spaced bins. The second step uses the frequency counts across the bins as a dependent variable in a local linear regression. Intuitively, the test looks for the presence of a discontinuity at the threshold in the density of the forcing variable. Unfortunately, this test is informative only if manipulation is monotonic. If the treatment induces some agents to manipulate the forcing variable in one direction and some agents in the other direction, the density may still appear continuous at the threshold, despite the manipulation. Additionally, manipulation may still be independent of potential

outcomes, so that this test does not obviate the need for a clear understanding and discussion of the relevant institutional details and incentives.

5.5.2 Balancing Tests and Covariates

Recall the implication of the local continuity assumption. Agents close to but on different sides of the threshold should have similar potential outcomes. Equivalently, these agents should be comparable both in terms of observable and unobservable characteristics. This suggests testing for balance (i.e., similarity) among the observable characteristics. There are several ways to go about executing these tests.

One could perform a visual analysis similar to that performed for the outcome variable. Specifically, create a number of nonoverlapping bins for the forcing variable, making sure that no bin contains points from both above and below the threshold. For each bin, plot the average characteristic over the midpoint for that bin. The average characteristic for the bins close to the cutoff should be similar on both sides of the threshold if the two groups are comparable. Alternatively, one can simply repeat the RDD analysis by replacing the outcome variable with each characteristic. Unlike the outcome variable, which should exhibit a discontinuity at the threshold, each characteristic should have an estimated treatment effect statistically, and economically, indistinguishable from zero.

Unfortunately, these tests do not address potential discontinuities in unobservables. As such, they cannot guarantee the internal validity of a RDD. Similarly, evidence of a discontinuity in these tests does not necessarily invalidate an RDD (van der Klaauw, 2008). Such a discontinuity is only relevant if the observed characteristic is related to the outcome of interest, y . This caveat suggests another test that examines the sensitivity of the treatment effect estimate to the inclusion of covariates other than the forcing variable. If the local continuity assumption is satisfied, then including covariates should only influence the precision of the estimates by absorbing residual variation. In essence, this test proposes expanding the specifications in Eqn (45), for sharp RDD,

$$y_i = \alpha + \beta d_i + f(x_i - x') + d_i \cdot g(x_i - x') + h(Z_i) + \varepsilon_i,$$

and Eqns (49) and (50), for fuzzy RDD:

$$\begin{aligned} y_i &= \alpha + \beta d_i + f(x_i - x') + h_y(Z_i) + \varepsilon_i, \\ d_i &= \delta + \phi T_i + g(x_i - x') + h_d(Z_i) + \omega_i, \end{aligned}$$

where h , h_y , and h_d are continuous functions of an exogenous covariate vector, Z_i . For example, Chava and Roberts (2008) show that their treatment effect estimates are largely

unaffected by inclusion of additional linear controls for firm and period fixed effects, cash flow, firm size, and several other characteristics. Alternatively, one can regress the outcome variable on the vector of observable characteristics and repeat the RDD analysis using the residuals as the outcome variable, instead of the outcome variable itself (Lee, 2008).

5.5.3 Falsification Tests

There may be situations in which the treatment did not exist or groups for which the treatment does not apply, perhaps because of eligibility considerations. In this case, one can execute the RDD for this era or group in the hopes of showing no estimated treatment effect. This analysis could reinforce the assertion that the estimated effect is not due to a coincidental discontinuity or discontinuity in unobservables.

Similarly, Kane (2003) suggests testing whether the actual cutoff fits the data better than other nearby cutoffs. To do so, one can estimate the model for a series of cutoffs and plot the corresponding log-likelihood values. A spike in the log-likelihood at the actual cutoff relative to the alternative false cutoffs can alleviate concerns that the estimated relation is spurious. Alternatively, one could simply look at the estimated treatment effects for each cutoff. The estimate corresponding to the true cutoff should be significantly larger than those at the alternative cutoffs, all of which should be close to zero.

6. Matching Methods

Matching methods estimate the counterfactual outcomes of subjects by using the outcomes from a subsample of “similar” subjects from the other group (treatment or control). For example, suppose we want to estimate the effect of a diet plan on individuals’ weights. For each person that participated in the diet plan, we could find a “match,” or similar person that did not participate in the plan, and, vice versa for each person that did not participate in the plan. By similar, we mean similar along weight-relevant dimensions, such as weight before starting the diet, height, occupation, health, etc. The weight difference between a person that undertook the diet plan and his match that did not undertake the plan measures the effect of the diet plan for that person.

One can immediately think of extensions to this method, as well as concerns. For instance, instead of using just one match per subject, we could use several matches. We could also weight the matches as a function of the quality of the match. Of course, how to

measure similarity and along which dimensions one should match are central to the proper implementation of this method.

Perhaps more important is the recognition that matching methods do not rely on a clear source of exogenous variation for identification. This fact is important and distinguishes matching from the methods discussed in sections 3 through 5. Matching does alleviate some of the concerns associated with linear regression, as we make clear below, and can mitigate asymptotic biases arising from endogeneity or self-selection. As such, matching can provide a useful robustness test for regression based analysis. However, matching by itself is unlikely to solve an endogeneity problem since it relies crucially on the ability of the econometrician to observe all outcome relevant determinants. Smith and Todd (2005) put it most directly, “... matching does not represent a ‘magic bullet’ that solves the selection problem in every context.” (page 3).

The remainder of the sections follows closely the discussion in Imbens (2004), to which we refer the reader for more details and further references. Some examples of matching estimators used in corporate finance settings include: Villalonga (2004), Colak and Whited (2007), Hellman, Lindsey, and Puri (2008), and Lemmon and Roberts (2010).

6.1 Treatment Effects and Identification Assumptions

The first important assumption for the identification of treatment effects (i.e., ATE, ATT, ATU) is referred to as unconfoundedness:

$$(y(0), y(1)) \perp d | X. \tag{51}$$

This assumption says that the potential outcomes ($y(0)$ and $y(1)$) are statistically independent (\perp) of treatment assignment (d) conditional on the observable covariates, $X = (x_1, \dots, x_k)$.³⁷ In other words, assignment to treatment and control groups is as though it were random, conditional on the observable characteristics of the subjects.

This assumption is akin to a stronger version of the orthogonality assumption for regression (assumption 4 from section 2.1). Consider the linear regression model assuming a constant treatment effect β_1 ,

$$y = \beta_0 + \beta_1 d + \beta_2 x_1 + \dots + \beta_{k+1} x_k + u.$$

³⁷This assumption is also referred to as “ignorable treatment assignment” (Rosenbaum and Rubin, 1983), “conditional independence” (Lechner, 1999), and “selection on observables” (Barnow, Cain, and Goldberger, 1980). An equivalent expression of this assumption is that $Pr(d = 1 | y(0), y(1), X) = Pr(d = 1 | X)$.

Unconfoundedness is equivalent to statistical independence of d and u conditional on (x_1, \dots, x_k) , a stronger assumption than orthogonality or mean independence.

The second identifying assumption is referred to as overlap:

$$0 < Pr(d = 1|X) < 1.$$

This assumption says that for each value of the covariates, there is a positive probability of being in the treatment group and in the control group. To see the importance of this assumption, imagine if it did not hold for some value of X , say X' . Specifically, if $Pr(d = 1|X = X') = 1$, then there are no control subjects with a covariate vector equal to X' . Practically speaking, this means that there are no subjects available in the control group that are similar in terms of covariate values to the treatment subjects with covariates equal to X' . This makes estimation of the counterfactual problematic since there are no comparable control subjects. A similar argument holds when $Pr(d = 1|X = X') = 0$ so that there are no comparable treatment subjects to match with controls at $X = X'$.

Under unconfoundedness and overlap, we can use the matched control (treatment) subjects to estimate the unobserved counterfactual and recover the treatment effects of interest. Consider the ATE for a subpopulation with a certain $X = X'$.

$$\begin{aligned} ATE(X') &\equiv E [y(1) - y(0)|X = X'] \\ &= E [y(1) - y(0)|d = d', X = X'] \\ &= E [y|d = 1, X = X'] - E [y|d = 0, X = X'] \end{aligned}$$

The first equality follows from unconfoundedness, and the second from $y = dy(1) + (1-d)y(0)$. To estimate the expectations in the last expression requires data for both treatment and control subjects at $X = X'$. This requirement illustrates the importance of the overlap assumption. To recover the unconditional ATE, one merely need integrate over the covariate distribution X .

6.2 The Propensity Score

An important result due to Rosenbaum and Rubin (1983) is that if one is willing to assume unconfoundedness, then conditioning on the entire k -dimensional vector X is unnecessary. Instead, one can condition on the 1-dimensional propensity score, $ps(x)$, defined as the probability of receiving treatment conditional on the covariates,

$$ps(x) \equiv Pr(d = 1|X) = E(d|X).$$

Researchers should be familiar with the propensity score since it is often estimated using discrete choice models, such as a logit or probit. In other words, unconfoundedness (Eqn (51)) implies that the potential outcomes are independent of treatment assignment conditional on $ps(x)$.

For more intuition on this result, consider the regression model

$$y = \beta_0 + \beta_1 d + \beta_2 x_1 + \cdots + \beta_{k+1} x_k + u.$$

Omitting the controls (x_1, \dots, x_k) will lead to bias in the estimated treatment effect, $\hat{\beta}_1$. If one were instead to condition on the propensity score, one removes the correlation between (x_1, \dots, x_k) and d because $(x_1, \dots, x_k) \perp d | ps(x)$. So, omitting (x_1, \dots, x_k) after conditioning on the propensity score no longer leads to bias, though it may lead to inefficiency.

The importance of this result becomes evident when considering most applications in empirical corporate finance. If X contains two binary variables, then matching is straightforward. Observations would be grouped into four cells and, assuming each cell is populated with both treatment and control observations, each observation would have an exact match. In other words, each treatment observation would have at least one matched control observation, and vice versa, with identical covariates.

This type of example is rarely seen in empirical corporate finance. The dimensionality of X is typically large and frequently contains continuous variables. This high-dimensionality implies that exact matches for all observations are typically impossible. It may even be difficult to find close matches along some dimensions. As a result, a large burden is placed on the choice of weighting scheme or norm to account for differences in covariates. Matching on the propensity score reduces the dimensionality of the problem and alleviates concerns over the choice of weighting schemes.

6.3 Matching on Covariates and the Propensity Score

How can we actually compute these matching estimators in practice? Start with a sample of observations on outcomes, covariates, and assignment indicators (y_i, X_i, d_i) . As a reminder, y and d are univariate random variables representing the outcome and assignment indicator, respectively; X is a k -dimensional vector of random variables assumed to be unaffected by the treatment. Let $l_m(i)$ be the index such that

$$d_l \neq d_i, \text{ and} \\ \sum_{j|d_j \neq d_i} l(\|X_j - X_i\| \leq \|X_l - X_i\|) = m.$$

In words, if i is the observation of interest, then $l_m(i)$ is the index of the observation in the group—treatment or control—that i is not in (hence, $d_l \neq d_i$), and that is the m^{th} closest in terms of the distance measure based on the norm $\|\cdot\|$. To clarify this idea, consider the 4^{th} observation ($i = 4$) and assume that it is in the treatment group. The index $l_1(4)$ points to the observation in the control group that is closest ($m = 1$) to the 4^{th} observation in terms of the distance between their covariates. The index $l_2(4)$ points to the observation in the control group that is next closest ($m = 2$) to the 4^{th} observation. And so on.

Now define $L_M(i) = \{l_1(i), \dots, l_M(i)\}$ to be the set of indices for the first M matches to unit i . The estimated or imputed potential outcomes for observation i are:

$$\hat{y}_i(0) = \begin{cases} y_i & \text{if } d_i = 0 \\ \frac{1}{M} \sum_{j \in L_M(i)} y_j & \text{if } d_i = 1 \end{cases}$$

$$\hat{y}_i(1) = \begin{cases} \frac{1}{M} \sum_{j \in L_M(i)} y_j & \text{if } d_i = 0 \\ y_i & \text{if } d_i = 1 \end{cases}$$

When observation i is in the treatment group $d_i = 1$, there is no need to impute the potential outcome $y_i(1)$ because we observe this value in y_i . However, we do not observe $y_i(0)$, which we estimate as the average outcome of the M closest matches to observation i in the control group. The intuition is similar when observation i is in the control group.

With estimates of the potential outcomes, the matching estimator of the average treatment effect (ATE) is:

$$\frac{1}{N} \sum_{i=1}^N [\hat{y}_i(1) - \hat{y}_i(0)].$$

The matching estimator of the average treatment effect for the treated (ATT) is:

$$\frac{1}{N_1} \sum_{i:d_i=1} [y_i - \hat{y}_i(0)],$$

where N_1 is the number of treated observations. Finally, the matching estimator of the average treatment effect for the untreated (ATU) is:

$$\frac{1}{N_0} \sum_{i:d_i=0} [\hat{y}_i(1) - y_i]$$

where N_0 is the number of untreated (i.e., control) observations. Thus, the ATT and ATU are simply average differences over the subsample of observations that are treated or untreated, respectively.

Alternatively, instead of matching directly on all of the covariates X , one can just match on the propensity score. In other words, redefine $l_m(i)$ to be the index such that

$$d_l \neq d_i, \text{ and} \\ \sum_{j|d_j \neq d_i} l(|\widehat{ps}(X_j) - \widehat{ps}(X_i)| \leq |\widehat{ps}(X_l) - \widehat{ps}(X_i)|) = m$$

This form of matching is justified by the result of Rosenbaum and Rubin (1983) discussed above. Execution of this procedure follows immediately from the discussion of matching on covariates.

In sum, matching is fairly straightforward. For each observation, find the best matches from the other group and use them to estimate the counterfactual outcome for that observation.

6.4 Practical Considerations

This simple recipe obfuscates a number of practical issues to consider when implementing matching. Are the identifying assumptions likely met in the data? Which distance metric $\|\cdot\|$ should be used? How many matches should one use for each observation (i.e., what should M be)? Should one match with replacement or without? Which covariates X should be used to match? Should one find matches for just treatment observations, just control, or both?

6.4.1 Assessing Unconfoundedness and Overlap

The key identifying assumption behind matching, unconfoundedness, is untestable because the counterfactual outcome is not observable. The analogy with regression estimators is immediate; the orthogonality between covariates and errors is untestable because the errors are unobservable. While matching avoids the functional form restrictions imposed by regression, it does require knowledge and measurement of the relevant covariates X , much like regression. As such, if selection occurs on unobservables, then matching falls prey to the same endogeneity problems in regression that arise from omitted variables. From a practical standpoint, matching will not solve a fundamental endogeneity problem. However, it can offer a nice robustness test.

That said, one can conduct a number of falsification tests to help alleviate concerns over violation of the unconfoundedness assumption. Rosenbaum (1987) suggests estimating a

treatment effect in a situation where there should not be an effect, a task accomplished in the presence of multiple control groups. These tests and their intuition are exactly analogous to those found in our discussion of natural experiments.

One example can be found in Lemmon and Roberts (2010) who use propensity score matching in conjunction with difference-in-differences estimation to identify the effect of credit supply contractions on corporate behavior. One result they find is that the contraction in borrowing among speculative-grade firms associated with the collapse of the junk bond market and regulatory reform in the early 1990s was greater among those firms located in the northeast portion of the country.

The identification concern is that aggregate demand fell more sharply in the northeast relative to the rest of the country so that the relatively larger contraction in borrowing among speculative grade borrowers was due to declining demand, and not a contraction in supply. To exclude this alternative, the authors re-estimate their treatment effect on investment-grade firms and unrated firms. If the contraction was due to more rapidly declining investment opportunities in the Northeast, one might expect to see a similar treatment effect among these other firms. The authors find no such effect among these other control groups.

The other identifying assumption is overlap. One way to inspect this assumption is to plot the distributions of covariates by treatment group. In one or two dimensions, this is straightforward. In higher dimensions, one can look at pairs of marginal distributions. However, this comparison may be uninformative about overlap because the assumption is about the joint, not marginal, distribution of the covariates.

Alternatively, one can inspect the quality of the worst matches. For each variable x_k of X , one can examine

$$\max_i |x_{ik} - X_{t_1(i),k}|. \quad (52)$$

This expression is the maximum over all observations of the matching discrepancy for component k of X . If this difference is large relative to the standard deviation of the x_k , then one might be concerned about the quality of the match.

For propensity score matching, one can inspect the distribution of propensity scores in treatment and control groups. If estimating the propensity score nonparametrically, then one may wish to undersmooth by choosing a bandwidth smaller than optimal or by including higher-order terms in a series expansion. Doing so may introduce noise but at the benefit of reduced bias.

There are several options for addressing a lack of overlap. One is to simply discard bad matches, or accept only matches with a propensity score difference below a certain threshold.

Likewise, one can drop all matches where individual covariates are severely mismatched using Eqn (52). One can also discard all treatment or control observations with estimated propensity scores above or below a certain value. What determines a “bad match” or how to choose the propensity score threshold is ultimately subjective, but requires some justification.

6.4.2 Choice of Distance Metric

When matching on covariates, there are several options for the distance metric. A starting point is the standard Euclidean metric:

$$\|X_i - X_j\| = \sqrt{(X_i - X_j)'(X_i - X_j)}$$

One drawback of this metric is its ignorance of variable scale. In practice, the covariates are standardized in one way or another. Abadie and Imbens (2006) suggest using the inverse of the covariates’ variances:

$$\|X_i - X_j\| = \sqrt{(X_i - X_j)' \text{diag}(\Sigma_X^{-1})(X_i - X_j)}$$

where Σ_X is the covariance matrix of the covariates, and $\text{diag}(\Sigma_X^{-1})$ is a diagonal matrix equal to the diagonal elements of Σ_X^{-1} and zero everywhere else. The most popular metric in practice is the Mahalanobis metric:

$$\|X_i - X_j\| = \sqrt{(X_i - X_j)' \Sigma_X^{-1} (X_i - X_j)},$$

which will reduce differences in covariates within matched pairs in all directions.³⁸

6.4.3 How to Estimate the Propensity Score?

As noted above, modeling of propensity scores is not new to most researchers in empirical corporate finance. However, the goal of modeling the propensity score is different. In particular, we are no longer interested in the sign, magnitude, and significance of a particular covariate. Rather, we are interested in estimating the propensity score as precisely as possible to eliminate, or at least mitigate, any selection bias in our estimate of the treatment effect.

There are a number of strategies for estimating the propensity score including: ordinary least squares, maximum likelihood (e.g., logit, probit), or a nonparametric approach, such

³⁸See footnote 6 of Imbens (2004) for an example in which the Mahalanobis metric can have unintended consequences. See Rubin and Thomas (1992) for a formal treatment of these distance metrics. See Zhao (2004) for an analysis of alternative metrics.

as a kernel estimator, series estimator, sieve estimator, etc.). Hirano, Imbens, and Ridder (2003) suggest the use of a nonparametric series estimator. The key considerations in the choice of estimator are accuracy and robustness. Practically speaking, it may be worth examining the robustness of one's results to several estimates of the propensity score.

6.4.4 How Many Matches?

We know of no objective rule for the optimal number of matches. Using a single (i.e., best) match leads to the least biased and most credible estimates, but also the least precise estimates. This tension reflects the usual bias-variance tradeoff in estimation. Thus, the goal should be to choose as many matches as possible, without sacrificing too much in terms of accuracy of the matches. Exactly what is too much is not well defined—any choice made by the researcher will have to be justified.

Dehejia and Wahba (2002) and Smith and Todd (2005) suggest several alternatives for choosing matches. Nearest neighbor matching simply chooses the m matches that are closest, as defined by the choice of distance metric. Alternatively, one can use caliper matching, in which all comparison observations falling within a defined radius of the relevant observation are chosen as matches. For example, when matching on the propensity score, one could choose all matches within $\pm 1\%$. An attractive property of caliper matching is that it relies on all matches falling within the caliper. This permits variation in the number of matched observations as a function of the quality of the match. For some observations, there will be many matches, for others few, all determined by the quality of the match.

In practice, it may be a good idea to examine variation in the estimated treatment effect for several different choices of the number of matches or caliper radii. If bias is a relevant concern among the choices, then one would expect to see variation in the estimated effect. If bias is not a concern, then the magnitude of the estimated effect should not vary much, though the precision (i.e., standard errors) may vary.

6.4.5 Match with or without Replacement?

Should one match with or without replacement? Matching with replacement means that each matching observation may be used more than once. This could happen if a particular control observation is a good match for two distinct treatment observations, for example. Matching with replacement allows for better matches and less bias, but at the expense of precision. Matching with replacement also has lower data requirements since observations

can be used multiple times. Finally, matching without replacement may lead to situations in which the estimated effect is sensitive to the order in which the treatment observations are matched (Rosenbaum, 1995).

We prefer to match with replacement since the primary objective of most empirical corporate finance studies is proper identification. Additionally, many studies have large amounts of data at their disposal, suggesting that statistical power is less of a concern.

6.4.6 Which Covariates?

The choice of covariates is obviously dictated by the particular phenomenon under study. However, some general rules apply when selecting covariates. First, variables that are affected by the treatment should *not* be included in the set of covariates X . Examples are other outcome variables or intermediate outcomes.

Reconsider the study of Lemmon and Roberts (2010). One of their experiments considers the relative behavior of speculative-grade rated firms in the Northeast (treatment) and speculative-grade rated firms elsewhere in the country (control). The treatment group consists of firms located in the Northeast and the outcomes of interest are financing and investment policy variables. Among their set of matching variables are firm characteristics and growth rates of outcome variables, which are used to ensure pre-treatment trend similarities. All of their matching variables are measured *prior* to the treatment in order to ensure that the matching variables are unaffected by the treatment.

Another general guideline suggested by Heckman et al., (1998) is that in order for matching estimators to have low bias, a rich set of variables related to treatment assignment and outcomes is needed. This is unsurprising. Identification of the treatment effects turns crucially on the ability to absorb all outcome relevant heterogeneity with observable measures.

6.4.7 Matches for Whom?

The treatment effect of interest will typically determine for which observations matches are needed. If interest lies in the ATE, then estimates of the counterfactuals for both treatment and control observations are needed. Thus, one need find matches for both observations in both groups. If one is interested only in the ATT, then we need only find matches for the treatment observations, and vice versa for the ATU. In many applications, emphasis is on the ATT, particularly program evaluation which is targeted toward a certain subset of the population. In this case, a deep pool of control observations relative to the pool of treatment observations is most relevant for estimation.

7. Panel Data Methods

Although a thorough treatment of panel data techniques is beyond the scope of this chapter, it is worth mentioning what these techniques actually accomplish in applied settings in corporate finance. As explained in Section 2.1.1, one of the most common causes of endogeneity in empirical corporate finance is omitted variables, and omitted variables are a problem because of the considerable heterogeneity present in many empirical corporate finance settings. Panel data can sometimes offer a partial, but by no means complete and costless, solution to this problem.

7.1 Fixed and Random Effects

We start with a simplified and sample version of Eqn (1) that contains only one regressor but in which we explicitly indicate the time and individual subscripts on the variables,

$$y_{it} = \beta_0 + \beta_1 x_{it} + u_{it}, \quad (i = 1, \dots, N; t = 1, \dots, T) \quad (53)$$

where the error term, u_{it} , can be decomposed as

$$u_{it} = c_i + e_{it}.$$

The term c_i can be interpreted as capturing the aggregate effect of all of the unobservable, time-invariant explanatory variables for y_{it} . To focus attention on the issues specific to panel data, we assume that e_{it} has a zero mean conditional on x_{it} and c_i for all t .

The relevant issue from an estimation perspective is whether c_i and x_{it} are correlated. If they are, then c_i is referred to as a “fixed effect.” If they are not, then c_i is referred to as a “random effect.” In the former case, endogeneity is obviously a concern since the explanatory variable is correlated with a component of the error term. In the latter, endogeneity is not a concern; however, the computation of standard errors is affected.³⁹

Two possible remedies to the endogeneity problem in the case of fixed effects is to run what is called a least squares dummy variable regression, which is simply the inclusion of firm-specific intercepts in Eqn (53). However, in many moderately large data sets, this approach is infeasible, so the usual and equivalent remedy is to apply OLS to the following deviations-from-individual-means regression:

³⁹Feasible Generalized Least Squares is often employed to estimate parameters in random effects situations.

$$\left(y_{it} - \frac{1}{T} \sum_{t=1}^T y_{it} \right) = \beta_1 \left(x_{it} - \frac{1}{T} \sum_{t=1}^T x_{it} \right) + \left(e_{it} - \frac{1}{T} \sum_{t=1}^T e_{it} \right). \quad (54)$$

The regression Eqn (54) does not contain the fixed effect, c_i , because $\left(c_i - T^{-1} \sum_{t=1}^T c_i \right) = 0$, so this transformation solves this particular endogeneity problem. Alternatively, one can remove the fixed effects through differencing and estimating the resulting equation by OLS

$$\Delta y_{it} = \beta_1 \Delta x_{it} + \Delta e_{it}.$$

Why might fixed effects arise? In regressions aimed at understanding managerial or employee behavior, any time-invariant individual characteristic that cannot be observed in the data at hand, such as education level, could contribute to the presence of a fixed effect. In regressions aimed at understanding firm behavior, specific sources of fixed effects depend on the application. In capital structure regressions, for example, a fixed effect might be related to unobservable technological differences across firms. In general, a fixed effect can capture any low frequency, unobservable explanatory variable, and this tendency is stronger when the regression has low explanatory power in the first place—a common situation in corporate finance.

Should a researcher always run Eqn (54) instead of Eqn (53) if panel data are available? The answer is not obvious. First, one should always try both specifications and check for statistical significance with a standard Hausman test in which the null is random effects and the alternative is fixed effects. However, one should also check to see whether the inclusion of fixed effects changes the coefficient magnitudes in an economically meaningful way. The reason is that including fixed effects reduces efficiency. Therefore, even if a Hausman test rejects the null of random effects, if the economic significance is little changed, the qualitative inferences from using pooled OLS on Eqn (53) can still be valid.

Fixed effects should be used with caution for additional reasons. First, including fixed effects can exacerbate measurement problems (Griliches and Mairesse, 1995). Second, if the dependent variable is a first differenced variable, such as investment or the change in corporate cash balances, and if the fixed effect is related to the *level* of the dependent variable, then the fixed effect has already been differenced out of the regression, and using a fixed-effects specification reduces efficiency. In practice, for example, fixed effects rarely tend to make important qualitative differences on the coefficients in investment regressions (Erickson and Whited, 2012), because investment is (roughly) the first difference of the capital stock. However, fixed effects do make important differences in the estimated coefficients in leverage

regressions (Lemmon, Roberts, and Zender, 2008), because leverage is a level and not a change.

Second, if the research question is inherently aimed at understanding cross-sectional variation in a variable, then fixed effects defeat this purpose. In the regression Eqn (54) all variables are forced to have the same mean (of zero). Therefore, the data variation that identifies β_1 is within-firm variation, and not the cross-sectional variation that is of interest. For example, Gan (2007) examines the effect on investment of land value changes in Japan in the early 1990s. The identifying data information is sharp cross sectional differences in the fall in land values for different firms. In this setting, including fixed effects would force all firms to have the same change in land values and would eliminate the data variation of interest. On the other hand, Khwaja and Mian (2008) specifically rely on firm fixed effects in order to identify the transmission of bank liquidity shocks onto borrowers' behaviors.

Third, suppose the explanatory variable is a lagged dependent variable, $y_{i,t-1}$. In this case the deviations-from-means transformation in Eqn (54) removes the fixed effect, but it induces a correlation between the error term $\left(e_{it} - T^{-1} \sum_{t=1}^T e_{it}\right)$ and $y_{i,t-1}$ because this composite error contains the term $e_{i,t-1}$.

In conclusion, fixed effects can ameliorate endogeneity concerns, but, as is the case with all econometric techniques, they should be used only after thinking carefully about the economic forces that might cause fixed effects to be an issue in the first place. Relatedly, fixed effects cannot remedy any arbitrary endogeneity problem and are by no means an endogeneity panacea. Indeed, they do nothing to address endogeneity associated with correlation between x_{it} and e_{it} . Further, in some instances fixed effects eliminate the most interesting or important variation researchers wish to explain. Examples in which fixed effects play a prominent role in identification include Khwaja and Mian (2008) and Hortacsu et al. (2010).

8. Econometric Solutions to Measurement Error

The use of proxy variables is widespread in empirical corporate finance, and the popularity of proxies is understandable, given that a great deal of corporate finance theory is couched in terms of inherently unobservable variables, such as investment opportunities or managerial perk consumption. In attempts to test these theories, most empirical studies therefore use observable variables as substitutes for these unobservable and sometimes nebulously defined quantities.

One obvious, but often costly, approach to addressing the proxy problem is to find better measures. Indeed, there are a number of papers that do exactly that. Graham (1996a,

1996b) simulates marginal tax rates in order to quantify the tax benefits of debt. Benmelech, Garmaise, and Moskowitz (2005) use information from commercial loan contracts to assess the importance of liquidation values on debt capacity. Benmelech (2009) uses detailed information on rail stock to better measure asset salability and its role in capital structure. However, despite these significant improvements, measurement error still persists.

It is worth asking why researchers should care, and whether proxies provide roughly the same inference as true unobservable variables. On one level, measurement error (the discrepancy between a proxy and its unobserved counterpart) is not a problem if all that one wants to say is that some observable proxy variable is correlated with another observable variable. For example, most leverage regressions typically yield a positive coefficient on the ratio of fixed assets to total assets. However, the more interesting questions relate to why firms with highly tangible assets (proxied by the ratio of fixed to total assets) have higher leverage. Once we start interpreting proxies as measures of some interesting economic concept, such as tangibility, then studies using these proxies become inherently more interesting, but all of the biases described in Section 2 become potential problems.

In this section, we outline both formal econometric techniques to deal with measurement error and informal but useful diagnostics to determine whether measurement error is a problem. We conclude with a discussion of strategies to avoid the use of proxies and how to use proxies when their use is unavoidable.

8.1 Instrumental Variables

For simplicity, we consider a version of the basic linear regression Eqn (1) that has only one explanatory variable:

$$y = \beta_0 + \beta_1 x^* + u. \quad (55)$$

We assume that the error term is uncorrelated with the regressors. Instead of observing x^* , we observe

$$x = x^* + w, \quad (56)$$

where w is uncorrelated with x^* . Suppose that one can find an instrument, z , that (i) is correlated with x^* (instrument quality), (ii) is uncorrelated with w (instrument validity), and (iii) is uncorrelated with u . This last condition intuitively means that z only affects y through its correlation with x^* . The IV estimation is straightforward, and can even be done in nonlinear regressions by replacing (ii) with an independence assumption (Hausman, Ichimura, Newey, and Powell, 1991, and Hausman, Newey, and Powell, 1995).

While it is easy to find variables that satisfy the first condition, and while it is easy to find variables that satisfy the second and third conditions (any irrelevant variables will do), it is very difficult to find variables that satisfy all three conditions at once. Finding instruments for measurement error in corporate finance is more difficult than finding instruments for simultaneity problems. The reason is that economic intuition or formal models can be used to find instruments in the case of simultaneity, but in the case of measurement error, we often lack any intuition for why there exists a gap between proxies included in a regression and the variables or concepts they represent.

For example, it is extremely hard to find instruments for managerial entrenchment indices based on counting antitakeover provisions (Gompers, Ishii, Metrick, 2003 and Bebchuck, Cohen, and Farrell, 2009). Entrenchment is a nebulous concept, so it is hard to conceptualize the difference between entrenchment and any one antitakeover provision, much less an unweighted count of several. Another example is the use of the volatility of a company's stock as a proxy for asymmetric information, as in Fee, Hadlock, and Thomas (2006). A valid instrument for this proxy would have to be highly correlated with asymmetric information but uncorrelated with the gap between asymmetric information and stock market volatility.

Several authors, beginning with Griliches and Hausman (1986), have suggested using lagged mismeasured regressors as instruments for the mismeasured regressor. Intuitively, this type of instrument is valid only if the measurement error is serially uncorrelated. However, it is hard to think of credible economic assumptions that could justify these econometric assumptions. One has to have good information about how measurement is done in order to be able to say much about the serial correlation of errors. Further, in many instances it is easy to think of credible reasons that the measurement error might be serially correlated. For example, Erickson and Whited (2000) discuss several of the sources of possible measurement error in Tobin's q and point out that many of these sources imply serially correlated measurement errors. In this case, using lagged instruments is not innocuous. Erickson and Whited (2012) demonstrate that in the context of investment regressions, using lagged values of x_{it} as instruments can result in the same biased coefficients that OLS produces if the necessary serial correlation assumptions are violated. Further, the usual tests of overidentifying restrictions have low power to detect this bias.

One interesting but difficult to implement remedy is repeated measurements. Suppose we replace Eqn (56) above with two measurement equations

$$\begin{aligned}x_{11} &= x^* + w_1, \\x_{12} &= x^* + w_2,\end{aligned}$$

where w_1 and w_2 are each uncorrelated with x^* , and uncorrelated with each other. Then it is possible to use x_{12} as an instrument for x_{11} . We emphasize that this remedy is only available if the two measurements are uncorrelated, and that this type of situation rarely presents itself outside an experimental setting. So although there are many instances in corporate finance in which one can find multiple proxies for the same unobservable variable, because these proxies are often constructed in similar manners or come from similar thought processes, the measurement errors are unlikely to be uncorrelated.

8.2 High Order Moment Estimators

One measurement error remedy that has been used with some success in investment and cash flow studies is high order moment estimators. We outline this technique using a stripped-down variant of the classic errors-in-variables model in Eqns (55) and (56) in which we set the intercept to zero. It is straightforward to extend the following discussion to the case in which Eqn (55) contains an intercept and any number of perfectly measured regressors.

The following assumptions are necessary: (i) (u, w, x^*) are *i.i.d.*, (ii) u , w , and x^* have finite moments of every order, (iii) $E(u) = E(w) = 0$, (iv) u and w are distributed **independently** of each other and of x^* , and (v) $\beta \neq 0$ and x^* is non-normally distributed. Assumptions (i)–(iii) are standard, but Assumption (iv) is stronger than the usual conditions on zero correlations or zero conditional expectations but is standard in most nonlinear error-in-variables estimators (e.g. Schennach, 2004).

To see the importance of Assumption (v), we square Eqn (55), multiply the result by Eqn (56), and take expectations of both sides. We obtain

$$E(y^2x) = \beta^2 E(x^{*3}). \quad (57)$$

Analogously, if we square Eqn (56), multiply the result by (55), and take expectations of both sides, we obtain

$$E(yx^2) = \beta E(x^{*3}). \quad (58)$$

As shown in Geary (1942), if $\beta \neq 0$ and $E(x^{*3}) \neq 0$, dividing Eqn (57) by Eqn (58) produces a consistent estimator for β :

$$\hat{\beta} = E(y^2x) / E(yx^2). \quad (59)$$

The estimator given by Eqn (59) is a third-order moment estimator. Inspection of Eqn (59) shows that the assumptions $\beta \neq 0$ and $E(x^{*3}) \neq 0$ are necessary for identification because one cannot divide by zero.

The estimators in Erickson and Whited (2002) build and improve upon this old result by combining the information in the third-order moment conditions (57) and (58) with the information in second- through seventh-order moment equations via GMM. High order moments cannot be estimated with as much precision as the second order moments on which conventional regression analysis is based. It is therefore important that the high order moment information be used as efficiently as possible. The third order moment estimator given by Eqn (59), although consistent, does not necessarily perform well in finite samples (Erickson and Whited, 2000, 2002), but the GMM estimators that combine information in many orders of moments can perform well on simulated data that resembles firm-level data on investment and Tobin's q .

One particular advantage of this technique is its potential usefulness. The GMM test of the overidentifying restrictions of the model has reasonable power to detect many types of misspecification that might plague regressions, such as heteroskedasticity and nonlinearity. A second useful feature is that it is possible to estimate the R^2 of Eqn (56), which can be expressed as $E(x^{*2}) / (E(x^{*2}) + E(w^2))$. As explained in section 2.1.3, this quantity is a useful index of proxy quality.

The economic intuition behind the Erickson and Whited estimators is easiest to see by observing that the estimator given by Eqn (59) has an instrumental variables interpretation. Recall that the problem for OLS in the classical errors-in-variable model can be shown by using Eqns (55) and (56) to write the relationship between the observable variables as

$$\hat{y} = \beta_0 + \beta x + (u - \beta w), \quad (60)$$

and then noting that x and the composite error $u - \beta w$ are correlated because they both depend on w . A valid instrument is one that satisfies the exclusion restriction that it not be correlated with $u - \beta w$ and that also satisfies the condition that it be highly correlated with x . The instrument $z = yx$ leads to exactly the same estimator as Eqn (59), and economic reasoning can be used as it would for any proposed instrument to verify whether it satisfies the conditions for instrument validity and relevance.

Straightforward algebra shows that the exclusion restrictions holds if the assumptions of the classical errors-in-variables model also hold. Therefore, using this instrument requires understanding the economic underpinnings of these assumptions. For example, Erickson and Whited (2000) provide a discussion of the assumptions necessary to apply high order moment estimators to investment regressions.

Straightforward algebra also shows that the condition for instrument relevance hinges on the assumption that x^* be skewed. The technique therefore works well when the mismeasured

regressor in question is marginal q . True marginal q is a shadow value, cannot therefore be negative, and must therefore be skewed. In addition, several models, such as Abel (1983) imply that marginal q , like many valuation ratios in finance, is highly skewed. Therefore, although this technique has been used successfully in studies that use an observable measure of Tobin's q for marginal q , (e.g. Erickson and Whited, 2000; Chen and Chen, 2012, and Riddick and Whited, 2009), it is by no means universally applicable.

8.3 Reverse Regression Bounds

One of the central themes of this section is that it is very difficult in most cases to find econometric remedies for measurement error. This situation therefore begs for diagnostics to determine whether measurement error might be biasing a coefficient of interest in a regression. One such useful diagnostic is reverse regressions bounds. We consider the case of one mismeasured regressor, x_1^* , and one perfectly measured regressor, x_2 . The regression and measurement equations are

$$y = \beta_0 + \beta_1 x_1^* + \beta_2 x_2 + u \quad (61)$$

$$x_1 = x_1^* + w, \quad (62)$$

where, without loss of generality, $\beta_1 > 0$ and $\beta_2 > 0$. Now, suppose we use the noisy proxy, x_1 , in place of x_1^* in Eqn (61) and then run the following two OLS regressions:

$$y = x_1 b_1 + x_2 b_2 + u \quad (63)$$

$$x_1 = y \left(\frac{1}{b_1^r} \right) + x_2 \left(\frac{-b_2^r}{b_1^r} \right) + u \left(\frac{-1}{b_1^r} \right). \quad (64)$$

Gini (1921) showed that the true coefficients (β_1, β_2) must lie, respectively, in the two intervals: (b_1, b_1^r) and (b_2, b_2^r) . To estimate the reverse regression coefficients, simply estimate Eqn (64) as a linear regression of x_1 on y and x_2 , use the estimated parameters to solve for b_1^r and b_2^r , and use the delta method to calculate standard errors.

This diagnostic is sometimes but not always informative. A useful example is in Erickson and Whited (2005), who examine a standard leverage regression that contains the market-to-book ratio, the ratio of fixed to total assets (tangibility), the log of sales, and earnings before interest and taxes (EBIT). The direct and reverse regression results are as follows.

	Market-to-book	Tangibility	Log sales	EBIT	R^2
Direct regression	-0.070 (0.003)	0.268 (0.012)	0.026 (0.001)	-0.138 (0.023)	0.216
Reverse regression	-0.738 (0.027)	-0.326 (0.039)	0.021 (0.005)	2.182 (0.133)	0.247

The market-to-book ratio is mismeasured in its role as a proxy for true investment opportunities, and suppose for now that it is the only mismeasured variable in this regression. Then the attenuation bias on the coefficient on market-to-book can be seen in the result that the true coefficient must lie in the interval $(-0.07, -0.74)$ and therefore be greater in absolute value than the OLS estimate. Note that the coefficient on log sales appears to be bounded above and below very tightly. This result occurs because log sales is only weakly correlated with market-to-book, and measurement error in one regressor affects the other regressors through its covariances with the other regressors. If a researcher finds the coefficient on log sales particularly interesting, then the reverse regression bound is a useful tool.

On the other hand, if the researcher is more interested in the coefficients on tangibility or EBIT, then the reverse regression intervals contain zero, so that this exercise cannot be used to determine the signs of these coefficients. It is still possible to obtain information about the signs of these coefficients by conducting the following thought experiment. Suppose we measure the proxy quality as the R^2 of Eqn (56). Then it is interesting to ask how low this proxy quality can be before the OLS estimate of a coefficient differs in sign from the true coefficient. Straightforward algebra shows that this threshold can be calculated as

$$R_x^2 + \hat{b}_1 \frac{\phi_x}{\phi_y} (1 - R_x^2),$$

where R_x^2 is the R^2 from regressing the mismeasured regressors on the other regressors, \hat{b}_1 is the OLS estimate of β_1 , ϕ_x is the coefficient on the perfectly measured regressor of interest (say EBIT) in a regression of the mismeasured regressor on all of the perfectly measured regressors, and ϕ_y is the coefficient on the perfectly measured regressor of interest in a regression of the dependent variable on all of the perfectly measured regressors. This type of threshold is useful either when it is near zero, which implies that the OLS estimate is likely giving the correct coefficient sign or when it is near one, which implies that the OLS estimate is almost certainly not delivering the correct coefficient sign. Erickson and Whited (2005) estimate these thresholds for tangibility and EBIT as 0.33 and 0.70. The second of these two thresholds implies that the measurement quality of the market-to-book ratio must be very high in order to infer a negative coefficient value. However, both of these thresholds

are difficult to interpret because neither is near and endpoint of the $(0, 1)$ interval, and there is limited empirical evidence on the measurement quality of market-to-book in its role as a proxy for investment opportunities.⁴⁰

8.4 Avoiding Proxies and Using Proxies Wisely

The key point of this section is that measurement error in proxies is difficult to deal with in most applications. So what is a researcher to do? We offer three suggestions.

- If one finds plausible the assumptions necessary to use one of the measurement error remedies described above, then use that remedy. Do not, however, blindly use remedies without thinking about the required assumptions. In particular, we recommend thinking very carefully about using lagged mismeasured regressors as instruments.
- Second, use proxies in such a way that their use makes it more difficult, rather than easier, to reject a null. In this way, one is more likely to commit a type II error than a type I error. Attenuation bias on a mismeasured regressor provides a clear example of this kind of approach if one is actually interested in the coefficient on the mismeasured regressor. If one is interested in using this approach, it is important, however, to remember that attenuation bias affects only the coefficient on a mismeasured regressor in a regression with one mismeasured regressor. The coefficients on other regressors can be biased in either direction, and in the case of multiple mismeasured regressors, the direction of coefficient bias is exceedingly hard to determine.
- Third, use the crude information in proxies in a crude way, that is, use a proxy to compare observations in either tail of the distribution of the proxy. In this case, even if the proxy conveys only noisy information about some underlying true variable, at least the observations in one tail of this distribution should be reliably different from those in the other.

9. Conclusion

This survey has provided a thorough and intuitive introduction to the latest econometric techniques designed to address endogeneity and identification concerns. Using examples from corporate finance, we have illustrated the practice and pitfalls associated with implementing

⁴⁰Whited (2001) estimates a measurement quality index of approximately 0.2.

these techniques. However, it is worth reemphasizing a message that is relevant for all techniques. There is no magic in econometrics. This notion is perhaps best summarized by the famous statistician, David Freedman, in his paper “Statistical Models and Shoe Leather” (Freedman, 1991):

I do not think that regression can carry much of the burden in a causal argument. Nor do regression equations, by themselves, give much help in controlling for confounding variables. Arguments based on statistical significance of coefficients seem generally suspect; so do causal interpretations of coefficients. More recent developments . . . may be quite interesting. However, technical fixes do not solve the problems, which are at a deeper level. (Page 292)

Though Freedman refers to regression, his arguments are just as applicable to some of the non-regression based methods discussed here (e.g., matching). As Freedman notes, statistical technique is rarely a substitute for good empirical design, high-quality data, and careful testing of empirical predictions against reality in a variety of settings. We hope that researchers will employ the tools discussed in this survey with these thoughts in mind.

Related, we recognize the fundamental subjectivity associated with addressing endogeneity concerns. Outside of controlled experiments, there is no way to guarantee that endogeneity problems are eliminated or sufficiently mitigated to ensure proper inferences. Ultimately, appropriately addressing endogeneity rests on the strength of one’s arguments supporting the identification strategy. To this end, we have stressed the importance of clearly discussing the relevant endogeneity concern — its causes and consequences — and how the proposed identification strategy addresses this issue. Only with clear and compelling arguments and analysis can one overcome endogeneity problems in observational studies.

Finally, we do not want our discussion to dissuade researchers from undertaking more descriptive studies where addressing endogeneity concerns may not be a first-order consideration. Ultimately, researchers seek to understand the causal forces behind economic phenomena. For this purpose, appropriately addressing endogeneity concerns is crucial. However, a first step toward this goal is often an interesting correlation whose interpretation is not yet clear. Descriptive analysis of new data often plays an integral role at the start of a research programme that can lay the foundation for future work focused on identifying the causal interpretations of those correlations.

References

- Abel, Andy, 1983, Optimal Investment Under Uncertainty, *American Economic Review* 73, 228-233.
- Agrawal, Ashwini, 2009, The impact of investor protection law on corporate policy: Evidence from the Blue Sky Laws, Working Paper (New York University Stern School of Business, New York, NY).
- Abadie, Alberto and Guido Imbens, 2006, Large sample properties of matching estimators for average treatment effects, *Econometrica* 74, 235-267.
- Almeida, Heitor, Murillo Campello, Bruno Laranjeira, and Scott Weisbenner, 2012, Corporate debt maturity and the real effects of the 2007 credit crisis, *Critical Finance Review*, 1, 3-58.
- Angrist, Joshua and Guido Imbens, 1994, Identification and estimation of local average treatment effects, *Econometrica* 62, 467-476.
- Angrist, Joshua, Guido Imbens, and Don Rubin, 1996, Identification and causal effects using instrumental variables, *Journal of the American Statistical Association* 91, 444-455.
- Angrist, Joshua and Alan Krueger, 1999, Empirical strategies in labor economics, in O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics* Vol. 3, 1277-1366.
- Angrist, Joshua and Jorn-Steffen Pischke, 2009, *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ.
- Asker, John and Alexander Ljungqvist, 2010, Competition and the structure of vertical relationships in capital markets, *Journal of Political Economy*, 118, 599-647.
- Baker, Malcolm and Yuhai Xuan, 2010, Under new management: Equity issues and the attribution of past returns, Working Paper, Harvard Business School.
- Bakke Tor-Erik, Candace Jens and Toni M. Whited, in press, The real effects of market liquidity: Causal evidence from delisting, *Finance Research Letters*.
- Barnow, Burt, Glen Cain and Arthur Goldberger, 1980, Issues in the analysis of selectivity bias, in E. Stromsdorfer and G. Farkas (Eds.), *Evaluation Studies*, Vol. 5, 43-59.
- Bebchuck, Lucian, Alma Cohen and Allen Ferrell, 2009, What matters in corporate governance?, *Review of Financial Studies* 22, 783-827.

- Becker, Bo, 2007 Geographical segmentation of U.S. capital markets, *Journal of Financial Economics* 85, 151-178.
- Becker, Bo and Per Stromberg, 2010, Equity-debtholder conflicts and capital structure, Working Paper, Harvard Business School.
- Benmelech, Effi, Mark Garmaise, and Tobias Moskowitz, 2005, Do liquidation values affect financial contracts? Evidence from commercial loan contracts and zoning regulation, *Quarterly Journal of Economics* 120, 1121-1154.
- Benmelech, Effi, 2009, Asset salability and debt maturity: Evidence from 19th Century American Railroads, *Review of Financial Studies* 22, 1545-1583.
- Bennedsen, Morten, Kasper Nielsen, Francisco Perez-Gonzalez and Daniel Wolfenzon, 2007, Inside the family firm: The role of families in succession decisions and performance, *Quarterly Journal of Economics* 122, 647-691.
- Bertrand, Marianne and Sendhil Mullainathan, 2003, Enjoying the quiet life? Corporate governance and managerial preferences, *Journal of Political Economy* 111, 1043-1075.
- Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan, 2004, How much should we trust differences- in-differences estimates?, *Quarterly Journal of Economics* 119, 249-275.
- Bertrand, Marianne, Antoinette Schoar, and David Thesmar, 2007, Banking deregulation and industry structure: Evidence from the French banking reforms of 1985, *Journal of Finance* 62, 597-628.
- Bhagat, Sanjai and Richard Jeffries, 2005, *The Econometrics of Corporate Governance Studies*, MIT Press, Cambridge, MA.
- Black, Dan, Jose Galdo and Jeffrey Smith, 2007, Evaluating the worker profiling and reemployment services system using a regression-discontinuity approach, *American Economic Review* 97, 104-107.
- Black, Bernard and Woochan Kim, 2011, The Effect of Board Structure on Firm Value: A Multiple Identification Strategies Approach Using Korean Data, Working Paper, Northwestern University.
- Blanchard, Olivier Jean, Florencio Lopez-de-Silanes and Andrei Shleifer, 1994, What do firms do with cash windfalls?, *Journal of Financial Economics* 36, 337-360.

- Blundell, Richard and Stephen D. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics* 87, 115-143.
- Bond, Stephen D. and Costas Meghir, 1994, Dynamic Investment Models and the Firm's Financial Policy, *Review of Economic Studies* 61, 197-222.
- Cain, Glen, 1975, Regression and selection models to improve nonexperimental comparisons, in C. Bennett and A. Lumsdain (Eds.) *Evaluation and Experiment*, 297-317.
- Chaney, Thomas, David Sraer and David Thesmar, in press, The collateral channel: How real estate shocks affect corporate investment, *American Economic Review*.
- Chava, Sudheer and Michael Roberts, 2008, How does financing impact investment? The role of debt covenant violations, *Journal of Finance* 63, 2085-2121.
- Chen, Huafeng (Jason) and Shaojun Chen, 2012, Investment-cash flow sensitivity cannot be a good measure of financial constraints: Evidence from the time series, *Journal of Financial Economics*, 103, 2012.
- Chen, Susan and Wilbert van der Klaauw, 2008, The work disincentive effects of the disability insurance program in the 1990s, *Journal of Econometrics* 142, 757-784.
- Çolak, Gönül and Toni M. Whited, 2007, Spin-offs, eivestitures, and conglomerate investment, *Review of Financial Studies* 20, 557-595.
- Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi, 2010, Plausibly exogenous, *Review of Economics and Statistics*, 94, 260-272.
- Core, John, Wayne Guay, and David Larcker, 2008, The power of the pen and executive compensation, *Journal of Financial Economics* 88: 1-25.
- Cummins, Jason, Kevin Hassett, and Steven S. Oliner, 2006, Investment Behavior, Observable Expectations, and Internal Funds, *American Economic Review* 96, 796-810.
- Dehejia, Rajeev and Sadek Wahba, 2002, Propensity score-matching methods for nonexperimental causal studies, *Review of Economics and Statistics* 84, 151-161.
- Erickson, Timothy and Toni M. Whited, 2000, Measurement error and the relationship between investment and q , *Journal of Political Economy* 108, 1027-1057.
- Erickson, Timothy and Toni M. Whited, 2002, Two-Step GMM Estimation of the Errors-in-Variables Model using High-Order Moments, *Econometric Theory* 18, 776-799.

- Erickson, Timothy and Toni M. Whited, 2005, Proxy quality thresholds: Theory and applications, *Finance Research Letters* 2, 131-151.
- Erickson, Timothy and Toni M. Whited, 2012, Treating Measurement Error in Tobin's q , *Review of Financial Studies*, 25, 1286-1329.
- Fan, Jianqing, 1992, Design-adaptive nonparametric regression, *Journal of the American Statistical Association* 87, 998-1004.
- Fan, Jianqing and Irene Gijbels, 1996, *Local Polynomial Modelling and its Applications*. Chapman and Hall, London.
- Faulkender, Michael and Mitchell Petersen, 2010, Investment and capital constraints: Repatriations under the American Jobs Creation Act, Working Paper, Northwestern University.
- Fazzari, Steven, R. Glenn Hubbard and Bruce C. Petersen, 1988, Financing constraints and corporate investment, *Brookings Papers on Economic Activity* 19, 141-206.
- Fee, C. Edward, Charles Hadlock and Shawn Thomas, 2006, Corporate Equity Ownership and the Governance of Product Market Relationships, *Journal of Finance* 61, 1217-1252.
- Fischer, Edwin O., Robert Heinkel and Josef Zechner, 1989, Dynamic capital structure choice: theory and tests, *Journal of Finance* 44, 19-40.
- Flannery, Mark and Kasturi Rangan, 2006, Partial adjustment toward target capital structures, *Journal of Financial Economics* 79, 469-506.
- Frank, Murray and Vidhan Goyal, 2009, Capital structure decisions: Which factors are reliably important? *Financial Management* 38, 1-37.
- Freedman, David, 1991, Statistical models and shoe leather, *Sociological Methodology* 21, 291-313.
- Gabaix, Xavier and Augustin Landier, 2008, Why has CEO pay increased so much? *Quarterly Journal of Economics* 123, 49-100.
- Gan, Jie, 2007, The Real effects of asset market bubbles: Loan- and firm-level evidence of a lending channel, *Review of Financial Studies* 20, 1941-1973.
- Garvey, Gerald T. and Gordon R. Hanka, 1999, Capital structure and corporate control: The effect of antitakeover statutes on firm leverage, *Journal of Finance* 54, 519-546.

- Geary, R. C., 1942, Inherent relations between random variables, *Proceedings of the Royal Irish Academy A* 47, 63-76.
- Gini, Corrado, 1921, Sull'interpolazione di una retta quando i valori della variabile indipendente sono affetti da errori accidentali, *Metroeconomica* 1, 63-82.
- Giroud, Xavier, Holger Mueller, Alexander Stomper, and Arne Westerkamp, 2010, Snow and leverage, Working Paper, New York University.
- Gomes, Joao, 2001, Financing investment, *American Economic Review* 91, 1263-1285.
- Gompers, Paul, Joy Ishii and Andrew Metrick, 2003, Corporate governance and equity prices, *Quarterly Journal of Economics* 118, 107-155.
- Gormley, Todd and David Matsa, 2011, Growing out of trouble: Corporate responses to liability risk, *Review of Financial Studies*, 24, 2781-2821. .
- Graham, John R., 1996a, Debt and the marginal tax rate, *Journal of Financial Economics* 41, 41-73.
- Graham, John R., 1996b, Proxies for the corporate marginal tax rate, *Journal of Financial Economics* 42, 187-221.
- Graham, John R., Roni Michaely and Michael R. Roberts, 2003, Do price discreteness and transactions costs affect stock returns? Comparing ex-dividend pricing before and after decimalization, *Journal of Finance* 58, 2611-2635.
- Griliches, Zvi, and Jerry A. Hausman., 1986, Errors in variables in panel data, *Journal of Econometrics* 31, 93-118.
- Griliches, Zvi and Jacques Mairesse, 1995, Production functions: The search for identification, *NBER Working Paper 5067*.
- Guiso, Luigi, Paola Sapienza and Luigi Zingales, 2004, Does local financial development matter? *Quarterly Journal of Economics* 119, 929-969.
- Hahn, Jinyong and Hausman, Jerry, 2005, Instrumental variable estimation with valid and invalid instruments, *Annales d'Economie et de Statistique* 79/80, 25-58
- Hahn, Jinyong, Petra Todd and Wilbert van der Klaauw, 2001, Identification and estimation of treatment effects with a regression-discontinuity design, *Econometrica* 69, 201-209.

- Hansen, Lars Peter and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1269-1286.
- Hausman, Jerry, Hidehiko Ichimura, Whitney Newey and James Powell, 1991, Measurement errors in polynomial regression models, *Journal of Econometrics* 50, 271-295.
- Hausman, Jerry, Whitney Newey and James Powell, 1995, Nonlinear errors in variables: Estimation of some Engel curves, *Journal of Econometrics* 65, 205-233.
- Hayashi, Fumio, 1982, Tobin's marginal q and average q: A neoclassical interpretation, *Econometrica* 50, 213-224
- Heckman, James, 1997, Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations, *Journal of Human Resources* 32, 441-462.
- Heckman, James, Hideo Ichimura and Petra Todd, 1998, Matching as an econometric evaluation estimator, *Review of Economic Studies* 65, 261-294.
- Heckman, James, Hideo Ichimura, James Smith and Petra Todd, 1998, Characterizing selection bias using experimental data, *Econometrica* 66, 1017-1098.
- Heckman, James and Richard Robb Jr., 1985, Alternative methods for evaluating the impact of interventions: An overview, *Journal of Econometrics* 30, 239-267.
- Hellman, Thomas, Laura Lindsey, and Manju Puri, 2008, Building relationships early: Banks in venture capital, *Review of Financial Studies* 21, 513-541.
- Hertzberg, Andrew, Jose Liberti and Daniel Paravisini, 2010, Information and incentives inside the firm: Evidence from loan officer rotation, *Journal of Finance*, 65, 795-828.
- Hirano, Keisuke, Guido Imbens and Geert Ridder, 2003, Efficient estimation of average treatment effects using the estimated propensity score, *Econometrica* 71, 1161-1189.
- Holland, Paul, 1986, Statistics and causal inference, *Journal of the American Statistical Association* 81, 945-971.
- Hortacsu, Ali, Gregor Matvos, Chad Syverson, and Sriram Venkataraman, 2010, Are consumers affected by durable goods makers' financial distress? The case of Automakers, Working Paper, University of Chicago.
- Huang, Rongbing and Jay Ritter, 2009, Testing theories of capital structure and estimating the speed of adjustment, *Journal of Financial and Quantitative Analysis* 44, 237-271.

- Iliev, Peter and Ivo Welch, 2010, Reconciling estimates of the speed of adjustment of leverage ratios, Working Paper, Brown University.
- Imbens, Guido, 2004, Nonparametric estimation of average treatment effects under exogeneity, *Review of Economics and Statistics* 86, 4-29.
- Imbens, Guido and Thomas Lemieux, 2008, Regression discontinuity designs: A guide to practice, *Journal of Econometrics* 142, 615-635.
- Ivashina, Victoria, 2009, Asymmetric information effects on loan spreads, *Journal of Financial Economics* 92, 300-319.
- Jayaratne, Jith and Philip E. Strahan, 1996, The finance-growth nexus: Evidence from bank branch deregulation, *Quarterly Journal of Economics* 101, 639-670.
- Kane, Thomas J., 2003, A quasi-experimental estimate of the impact of financial aid on college-going, NBER Working Paper No., W9703.
- Keys, Benjamin, Tanmoy Mukherjee, Amit Seru and Vikrant Vig, 2010, Did securitization lead to lax screening, Working Paper (University of Chicago Booth School of Business, Chicago, IL).
- Klepper, Steven and Edward E. Leamer, 1984, Consistent Sets of Estimates for Regressions with Errors in All Variables, *Econometrica* 52, 163-183.
- Korajczyk, Robert A. and Amnon Levy, 2003, Capital structure choice: Macroeconomic conditions and financial constraints, *Journal of Financial Economics* 68, 75-109.
- Khwaja, Asim, and Atif Mian, 2008, Tracing the impact of bank liquidity shocks: Evidence from an emerging market, *American Economic Review* 98, 1413-1442.
- Leary, Mark T., 2009, Bank loan supply, lender choice, and corporate capital structure, *Journal of Finance* 64, 1143-1185.
- Lechner, Michael, 1999, Earnings and employment effects of continuous off-the-job training in East Germany after unification, *Journal of Business and Economic Statistics* 17, 74-90.
- Li, Kai and N.R. Prabhala, 2007, Self-selection models in corporate finance, in B. Espen Eckbo (ed.), *Handbook of Corporate Finance: Empirical Corporate Finance*, Elsevier, Amsterdam.

- Lee, David S., 2008, Randomized Experiments from Non-random Selection in U.S. House Elections, *Journal of Econometrics* 142, 675-697.
- Lee, David S. and Thomas Lemieux, 2010, Regression discontinuity designs in economics, *Journal of Economic Literature* 48, 281-355.
- Leland, Hayne, 1994, Corporate Debt Value, Bond Covenants, and Optimal Capital Structure, *Journal of Finance* 49, 1213-1252.
- Lemmon, Michael and Michael R. Roberts, 2010, The response of corporate financing investment to changes in the supply of credit, *Journal of Financial and Quantitative Analysis* 45, 555-587.
- Lemmon, Michael, Michael R. Roberts and Jaime F. Zender, 2008, Back to the beginning: Persistence and the cross-section of corporate capital structure, *Journal of Finance* 63, 1575-1608.
- Ludwig, Jens and Douglas L. Miller, 2007, Does Head Start improve children's life chances? Evidence from a regression discontinuity design, *Quarterly Journal of Economics* 122, 159-208.
- McCrary, Justin, 2008, Manipulation of the running variable in the regression discontinuity design: A density test, *Journal of Econometrics* 142, 698-714.
- Meyer, Bruce D., 1995, Natural and quasi-experiments in economics, *Journal of Business and Economic Statistics* 13, 151-161.
- Morse, Adair, 2011, Payday lenders: Heroes or villains? *Journal of Financial Economics*, 102, 28-44.
- Melzer, Brian, 2011, The real cost of credit access: Evidence from the payday lending market, *Quarterly Journal of Economics*, 126, 517-555.
- Murfin, Justin, 2010, The supply-side determinants of loan contract strictness, Working Paper, Yale University.
- Newey, Whitney K., James L. Powell, and James R. Walker, 1990, Semiparametric estimation of selection models: Some empirical results, *American Economic Review* 80, 324-328.
- Petersen, Mitchell A., 2009, Estimating standard errors in panel data sets, *Review of Financial Studies* 22, 435-480.

- Porter, Jack, 2003, Asymptotic bias and optimal convergence rates for semiparametric kernel estimators in the regression discontinuity model, Working Paper (University of Wisconsin Madison, Madison, WI).
- Rajan, Raghuram G. and Luigi Zingales, 1995, What do we know about capital structure? Some evidence from international data, *Journal of Finance* 50, 1421-1460.
- Riddick, Leigh A. and Toni M. Whited, 2009, The corporate propensity to save, *Journal of Finance* 64, 1729-1766.
- Roberts, Michael R. and Amir Sufi, 2009a, Control rights and capital structure: An empirical investigation, *Journal of Finance* 64, 1657-1695.
- Roberts, Michael R. and Amir Sufi, 2009b, Renegotiation of Financial Contracts: Evidence from Private Credit Agreements, *Journal of Financial Economics* 93, 159-184.
- Rosenbaum, Paul, 1987, The role of a second control group in an observational study, *Statistical Science* 2, 292-316.
- Rosenbaum, Paul, 1995, *Observational Studies*, Springer-Verlag, New York.
- Rosenbaum, Paul and Don Rubin, 1983, The central role of the propensity score in observational studies for causal effects, *Biometrika* 70, 41-55.
- Rubin, Don, 1974, Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology* 66, 688-701.
- Rubin, Don and Neal Thomas, 1992, Affinely invariant matching methods with ellipsoidal distributions, *Annals of Statistics* 20, 1079-1093.
- Schennach, Susanne M., 2004, Estimation of nonlinear models with measurement error, *Econometrica* 72, 33-75.
- Schnabl, Philipp, 2010, Financial globalization and the transmission of bank liquidity shocks: Evidence from an emerging market, Working paper.
- Schoar, Antoinette and Ebonya Washington, 2010, Are the seeds of bad governance sown in good times? Working Paper, MIT.
- Smith, Jeffrey and Petra Todd, 2005, Does matching address Lalonde's critique of nonexperimental estimators?, *Journal of Econometrics*, 305-353.

- Stock, James H., Jonathan H. Wright and Motohiro Yogo, 2002, A survey of weak instruments and weak identification in Generalized Method of Moments, *Journal of the American Statistical Association*, 20, 518-29.
- Stock, James H. and Mark W. Watson, 2007, Introduction to Econometrics. Addison-Wesley, Reading, MA.
- Stock, James H. and Motohiro Yogo, 2005, Testing for weak instruments in linear IV regression, Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, 80-108. Cambridge University Press, Cambridge.
- Sufi, Amir, 2009, The real effects of debt certification: Evidence from the introduction of bank loan ratings, *Review of Financial Studies* 22, 1659-1691.
- Tang, Tony, 2009, Information asymmetry and firms' credit market access: Evidence from Moody's credit rating format refinement, *Journal of Financial Economics* 93, 325-351.
- Titman, Sheridan and Roberto Wessels, 1988, The determinants of capital structure choice, *Journal of Finance* 43, 1-19.
- Tsoutsoura, Margarita, 2010, The effect of succession taxes on family firm investment: Evidence from a natural experiment, Working Paper (University of Chicago Booth School of Business, Chicago, IL).
- Van der Klaauw, Wilbert, 2002, Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach, *International Economic Review* 43, 1249-1287.
- Van der Klaauw, Wilbert, 2008, Breaking the link between poverty and low student achievement: An evaluation of Title I, *Journal of Econometrics* 142, 731-756.
- Villalonga, Belen, 2004, Does diversification cause the "Diversification Discount"? *Financial Management* Summer, 5-27.
- Whited, Toni M., 1992, Debt, liquidity constraints, and corporate investment: Evidence from panel data, *Journal of Finance* 47, 1425-1460.
- Whited, Toni M., 2001, Is it inefficient investment that causes the diversification discount?, *Journal of Finance* 56, 1667-1692.
- Wooldridge, Jeffrey M., 2002, *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

Zhao, Zhong, 2004, Using matching to estimate treatment effects: Data requirements, matching metrics, and Monte Carlo evidence, *Review of Economics and Statistics* 86, 91-107.