Journal of Accounting and Economics 49 (2010) 186–205

Contents lists available at ScienceDirect

Journal of Accounting and Economics

journal homepage: www.elsevier.com/locate/jae

On the use of instrumental variables in accounting research ☆

David F. Larcker [a,*], Tjomme O. Rusticus [b]

[a] Stanford Graduate School of Business, 518 Memorial Way, Stanford, CA 94305-5015, USA
[b] Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Room 6216, Evanston, IL 60208, USA

ARTICLE INFO

Article history:
Received 14 April 2005
Received in revised form
6 May 2009
Accepted 18 November 2009
Available online 26 November 2009

JEL classification:
C30
G30
M41
M43

Keywords:
Endogeneity
Instrumental variables
Disclosure
Cost of capital

ABSTRACT

Instrumental variable (IV) methods are commonly used in accounting research (e.g., earnings management, corporate governance, executive compensation, and disclosure research) when the regressor variables are endogenous. While IV estimation is the standard textbook solution to mitigating endogeneity problems, the appropriateness of IV methods in typical accounting research settings is not obvious. Drawing on recent advances in statistics and econometrics, we identify conditions under which IV methods are preferred to OLS estimates and propose a series of tests for research studies employing IV methods. We illustrate these ideas by examining the relation between corporate disclosure and the cost of capital.

1. Introduction

Instrumental variable (IV) methods are commonly used in accounting research to resolve econometric problems with observational data, such as when the outcome and explanatory variables are simultaneously determined (i.e., simultaneous-equation bias). Another problem occurs when a variable that affects both the outcome and explanatory variables is not included in the regression model (i.e., correlated omitted variable bias). Both of these problems frequently occur in accounting research. To resolve these problems, instrumental variable methods are used in both multiple equation models (to address simultaneity) and single-equation models (to address omitted variables).

In a typical IV application, the researcher first selects a set of variables that are assumed to be exogenous and then uses two-stage-least-squares (2SLS) or similar estimation methods to estimate the coefficients in the regression model. This standard textbook solution to endogeneity is appropriate if the researcher can find instrumental variables that are correlated with the endogenous regressor but uncorrelated with the error in the structural equation. However, as Maddala

☆ We appreciate helpful comments from Thomas Lys and Ross Watts (the editors), M. Laurentius Marais and Kin Lo (the referees), Christopher Armstrong, William Beaver, John Core, Theodore Goodman, Ian Gow, Robert Holthausen, Wayne Guay, Laurence van Lent, Christian Leuz, Russell Lundholm, Jeffrey Ng, Valeri Nikolaev, Peter Reiss, Scott Richardson, Catherine Schrand, Daniel Taylor, Rodrigo Verdi, Ran Wei, and workshop participants at the University of Pennsylvania. Rusticus is grateful for financial support from the Accounting Research Center at the Kellogg School of Management (Northwestern University) and the Deloitte & Touche Foundation.

* Corresponding author.
E-mail addresses: larcker_david@gsb.stanford.edu (D.F. Larcker), t-rusticus@northwestern.edu (T.O. Rusticus).

0165-4101/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jacceco.2009.11.004

(1977, p. 154) points out "Where do you get such a variable?" Similarly, Reiss and Wolak (2001) discuss the "magic" of finding an instrumental variable and cynically suggest that the best instrumental variable is developed by simply adding random error to the endogenous variable (which will be correlated with the original variable by construction). Consequently, it is necessary for researchers to understand the consequences of using instrumental variables that do not precisely conform to the necessary assumptions about these variables.

The purpose of this paper is to evaluate IV applications in accounting research, synthesize the extensive literature in statistics and econometrics on IV estimators, and provide accounting researchers with a framework to guide the use of IV methods. Our analytical results and numerical simulations indicate that when the instrument is only weakly correlated with the regressor, IV methods can produce highly biased estimates when the instrumental variable is even slightly endogenous. In those cases, it is likely that IV estimates are more biased and more likely to provide the wrong statistical inference than simple OLS estimates that make no correction for endogeneity.[1]

It is important to highlight that our analysis should *not* be interpreted as indicating that it is impossible for accounting researchers to address endogeneity, and thus that there is no reason to even make an attempt at dealing with endogeneity.[2] Rather, our analysis illustrates that researchers need to carefully justify their instrumental variables using economic theory and report contemporary specification tests for weak instruments and over-identifying restrictions. We believe that it is also useful to assess the sensitivity of OLS results to unobserved correlated (moderator and suppressor) variables using the methods developed by researchers such as Frank (2000), Rosenbaum (2002), and DiPrete and Gangl (2004).

The remainder of the paper consists of six sections. In Section 2, we examine IV applications in 42 papers published in *Journal of Accounting Research*, *Journal of Accounting and Economics*, or *The Accounting Review* during the time period from 1995 to 2005. In general, the variables selected as instruments seem largely arbitrary and are not justified by any rigorous theoretical discussion. Moreover, few diagnostic statistics are reported in the published articles and this makes it virtually impossible for the reader to assess the quality of the IV application.

In Sections 3 and 4, we discuss the asymptotic and finite sample properties of OLS and IV estimators. In this discussion, we focus on situations where the selected instrumental variables are not completely exogenous (i.e., the instruments that are correlated with the error term in the structural model or "semi-endogenous"), and problems with "weak" instruments (i.e., instruments that explain only a small proportion of the variation in the endogenous variable). These results are then used in Section 5 to develop a framework for accounting researchers using IV methods.

Section 6 compares the results produced by OLS and IV estimators in a contemporary accounting research setting where there is substantial reason to suspect that the primary regressor variable is endogenous. In particular, we examine the association between corporate disclosure and the cost of capital and conclude that in the context of our example OLS (which finds no statistical association) is preferred to IV estimation (which finds a negative statistical association). Although endogeneity remains a problem for the OLS results, the IV estimation is even less reliable than OLS. In Section 7, we illustrate the use of a methodology developed by Frank (2000) that enables the researcher to assess the possible impact of endogeneity on OLS parameter estimates. Concluding remarks about IV estimation in accounting research are provided in Section 8.

2. Instrumental variable applications in contemporary accounting research

In order to provide some insight into the use of IV estimation by accounting researchers, we conducted an electronic search for the terms "2SLS," "3SLS," "instrumental variable," and "endogeneity" for papers published in *Journal of Accounting Research*, *Journal of Accounting and Economics*, or *The Accounting Review* during the time period from 1995 to 2005. This search produced 42 articles that applied IV methods (listed in Table 1) in the study of earnings management, external disclosure, other financial accounting topics, managerial accounting, auditing, and corporate governance.

Accounting researchers generally use instrumental variables in an attempt to mitigate the biases caused by endogeneity of the predictor variables or to identify a system of simultaneous equations.[3] Of the 42 papers in our sample, 15 papers use IVs in a single-equation 2SLS, 7 in a Heckman-type model (e.g., Heckman, 1978, 1979), and 20 in a simultaneous-equation model (see Table 2, Panel A). IV methods are used in approximately equal proportion for the primary empirical results or robustness analyses (Table 2, Panel B). One unusual aspect of the typical robustness analysis is that the researchers frequently comment that their results are "robust" to endogeneity if the IV and OLS results produce similar

[1] Accounting researchers differ in what they mean by endogeneity. We follow the definition in econometrics (Wooldridge, 2002, p. 50) and label a variable endogenous if it is correlated with the error term. Common econometric problems such as simultaneity bias and omitted variables fall under this definition.

[2] Another unfortunate aspect of issues discussed in this paper concerns the role of endogeneity in the review process. We very much agree with the editorial by Shugan (2004) which deplores how endogeneity has become the (perhaps primary) reason for rejecting papers. Obviously, endogeneity is a serious problem, but there costs and benefits to various solutions to this econometric problem.

[3] Instrumental variables are also used to mitigate measurement error in the independent variables. This has a long history in economic research and applications in the accounting literature date back to at least Beaver et al. (1970). We do not include these applications in our discussion because sophisticated latent variable models exist to address measurement error issues when there are multiple indicators for the same construct.

Table 1
Accounting research articles that use instrumental variable methods.
The sample is based on an electronic search for the terms "2SLS," "3SLS," "instrumental variable," and "endogeneity" for papers published in *Journal of Accounting Research* (JAR), *Journal of Accounting and Economics* (JAE), or *The Accounting Review* (TAR) during the time period from 1995 to 2005.

Earnings Management (EM)
Anderson et al. (JAE, 2004)
Barton (TAR, 2001)
Beatty et al. (JAR, 1995)
Darrough and Rangan (JAR, 2005)
DeFond et al. (JAR, 2002)
D'Souza (TAR, 1998)
Haw et al. (JAR, 2004)
Hope (JAR, 2003)
Hunt et al. (JAE, 1996)
Kang and Sivaramakrishnan (JAR, 1995)

Disclosure (DIS)
Barton and Waymire (JAE, 2004)
Bushee et al. (JAE, 2003)
Kasznik (JAR, 1999)
Lang et al. (JAR, 2004)
Leuz and Verrecchia (JAR, 2000)

Auditing (AUD)
Copley et al. (JAR, 1995)
Khurana and Raman (TAR, 2004)
Weber and Willenborg (JAR, 2003)
Whisenant et al. (JAR, 2003)
Willenborg (JAR, 1999)

Other Financial Accounting (OTH)
Aboody (JAE, 1996)
Aboody et al. (TAR, 2004)
Ball and Shivakumar (JAE, 2005)
Barth et al. (JAR, 2001)
Beaver et al. (JAE, 1997)
Bell et al. (TAR, 2002)
Callen et al. (JAE 2005)
Frank (JAR, 2002)
Kothari and Zimmerman (JAE, 1995)
Lev and Sougiannis (JAE, 1996)
Loudder et al. (TAR, 1996)
Phillips (TAR, 2003)
Rajgopal et al. (JAR, 2003)
Shi (JAE, 2003)

Management Accounting/Compensation (MAC)
Abernethy et al. (TAR, 2004)
Hanlon et al. (JAE, 2003)
Holthausen et al. (JAE, 1995)
Keating (JAE, 1997)
Murphy (JAE, 2000)
Nagar (TAR, 2002)
Rajgopal and Shevlin (JAE, 2002)
Roulstone (JAR, 2003)

Table 2
Descriptive statistics for the accounting research articles that use instrumental variable methods.
The sample is based on an electronic search for the terms "2SLS," "3SLS," "instrumental variable," and "endogeneity" for papers published in *Journal of Accounting Research*, *Journal of Accounting and Economics*, or *The Accounting Review* during the time period from 1995 to 2005.

Panel A. Type of application	EM	DIS	AUD	OTH	MAC	Total
Recursive models						
Standard two-stage-least-squares	3	1	1	8	2	15
First stage is a probit model (Heckman)	0	2	2	2	1	7
Non-recursive models						
Simultaneous equations models	7	2	2	4	5	20

Panel B. Features of the application	**Two stage** # (%)	**Heckman** # (%)	**Simultaneous** # (%)	**Total** # (%)
Importance of instrumental variables				
Instrumental variables are used in main test	8 (53%)	3 (43%)	12 (60%)	23 (55%)
Instrumental variables are used as robustness test	7 (47%)	4 (57%)	8 (40%)	19 (45%)
Explanation in the paper				
Extensive discussion of model/method	5 (33%)	5 (71%)	16 (80%)	26 (62%)
Discussion of instruments	10 (67%)	1 (14%)	12 (60%)	23 (55%)
Justification of instruments	6 (40%)	0 (0%)	3 (15%)	9 (21%)
Reported empirical work				
Report first-stage coefficients	6 (40%)	5 (71%)	2 (10%)	13 (31%)
Report first-stage explanatory power	5 (33%)	4 (57%)	6 (30%)	15 (36%)
Report explanatory power of instruments	3 (20%)	0 (0%)	1 (5%)	4 (10%)
Standard regression (e.g., OLS) also reported	12 (80%)	6 (86%)	12 (60%)	30 (71%)
Reported tests				
Hausman test (or functional equivalent)	6 (40%)	4 (57%)	15 (75%)	25 (60%)
Over-identifying restrictions test	1 (7%)	0 (0%)	3 (15%)	4 (10%)
Number of papers by category	15	7	20	42

estimates. Unfortunately, if there are theoretical reasons to suspect serious endogeneity concerns, the similarity of the results may also indicate that the selected IVs are inadequate, as opposed to the reported results being unaffected by endogeneity.

In general, there is little attempt in the typical accounting study to develop a model that explicitly identifies and justifies the endogenous (or choice) variables and the exogenous and instrumental variables (i.e., those variables that are assumed to be either pre-determined or are outside the model being examined). Authors should articulate their theory in a manner that explains their selection of the variable, not select the variable and then justify it. The ideal approach is perhaps as follows: develop an economic theory of the decision making process regarding the relation of interest, translate the theory into a set of structural equation models that describe the decision setting, precisely identify the endogenous and exogenous variables, develop the reduced form equations where the endogenous variables are only functions of exogenous variables, and estimate the reduced form model parameters. Assuming that the model is identified, the structural equation parameters of interest can then be derived from the estimates obtained from the reduced form equations.

In our survey, it appears that most accounting researchers estimate some type of "convenient representation" that is assumed to be the appropriate reduced form of the underlying structural equation model. Moreover, there is almost no discussion regarding the choice of specific variables for instruments (only slightly more than half of the papers discuss the instruments at all). Researchers do not rigorously discuss why the variables selected as instruments are assumed to be exogenous (i.e., uncorrelated with the error term in the structural model) or why the instrumental variables exhibit a lower correlation with the structural equation error term than the endogenous regressor variable (nearly 80% of the papers provide no justification for the choice of instrument whatsoever).

One potential explanation for these observations is that better documentation was included in the original working paper versions or in the correspondence with the reviewers. For example, if a paper provides little justification for an instrument, perhaps it is because both the researcher(s) and the reviewer(s) believed the choice to be compelling. Similarly, the researchers may have added instrumental variable estimation as a sensitivity check in response to reviewer comments, without wanting to unduly lengthen the paper. However, we believe that in order for readers to be able to evaluate the quality of the instrumental variable estimation, researchers need to include, and reviewers should insist on, a thorough justification of the instruments along with a detailed empirical analysis.

As we discuss below, it is important to report the coefficient estimates and the explanatory power for the first-stage model. However, the results for the first-stage estimation are not generally reported in sufficient detail to judge the adequacy of the IV application. One problematic aspect of accounting IV applications is that if a first-stage R^2 is reported, it is the explanatory power for the *total* first-stage model, and not the *partial* explanatory power for the instruments that are unique to the first-stage regression.[4] Thus, using reported first-stage explanatory power would lead to a substantial overstatement of the strength of the instrumental variables in the first-stage regression. Finally, it is common for accounting researchers to report the classic Hausman (1978) test for endogeneity (or a test on the inverse Mill's ratio in a Heckman selection model). However, it is very uncommon for these studies to report the over-identifying restriction test (see Hausman, 1978 and Godfrey and Hutton, 1994) that should be conducted before implementing the Hausman test. The absence of the over-identifying restriction test makes it extremely difficult to assess the validity of IV application.

3. *Asymptotic* properties of instrumental variable estimators

3.1. General structure of the simultaneous-equation model

We start with the general linear model describing the relation between two or more variables of interest. In its basic form, such a simultaneous-equation model has two equations:

$$\begin{aligned} y &= \beta x + \sum_{k=1}^{K} \gamma_k x_k + u \\ x &= \delta y + \sum_{j=1}^{J} \theta_j x_j + \varepsilon \end{aligned} \tag{1}$$

In this system, x and y are the two endogenous variables of interest and the x_j's and x_k's are the control variables (assumed to be exogenous). This simultaneous-equation model captures a wide variety of economic research topics. For example, these equations could represent supply and demand functions, which simultaneously determine price and quantity (denoted by x and y). Alternatively, these could be two competing hypotheses on the direction of causality between x and y. As shown in Table 2, a substantial portion of the IV applications in the accounting literature consists of such simultaneous-equation models.

One way to estimate the system in Eq. (1) is to use OLS, and estimate each equation separately. However, since x is a function of y, and y is a function of u, the correlation between x and u will be non-zero, implying that the OLS estimates are inconsistent. Even when δ is zero, OLS is not necessarily consistent. For example when u and ε are correlated, this would

[4] The typical analysis in empirical research involves an endogenous y that is a function of an endogenous x-variable and a set of exogenous control variables (z_1). In addition, there are multiple instruments, exogenous variables (z_2) that are not included in the equation describing y. In this case, the proper measure of the strength of the instrument is the *partial* R^2. The partial R^2 can be easily computed using: $(R^2_{y,z}-R^2_{y,z_1})/(1-R^2_{y,z_1})$, where z is the combined set of z_1 and z_2.

also induce correlation between x and u resulting in biased coefficients. Only when δ is zero and x and u are uncorrelated will the OLS estimate of β be consistent.

The alternative estimation method is to use exclusion restrictions (i.e., instrumental variables). If there are variables in J (the set of x_j's) that are not in K (the set of x_k's), it is possible to estimate β using those variables in J as the instruments for x. Similarly, if there variables in K that are not in J, it is possible to estimate δ using those variables in K as the instruments for y. These instrumental variable estimates of β will be consistent if the instruments for x are uncorrelated with u. The instrumental variable estimate for β can be obtained using two-stage-least-squares (2SLS).

While the 2SLS estimate is consistent when the instruments are exogenous, it is generally not efficient (i.e., there are other instrumental variable estimators that asymptotically have lower standard errors). Approaches such as three-stage-least-squares (3SLS) and maximum likelihood estimation (MLE) utilize the correlation structure between u and ε to achieve greater efficiency. However, this greater efficiency comes at a cost. In these methods information from one equation is used in estimating other equations, which implies that for *any* equation in (1) to be estimated consistently *all* instruments in *all* equations in (1) need to be exogenous (e.g., Wooldridge, p. 199). In contrast, when using 2SLS, an equation can be consistently estimated as long as the instruments for that particular equation are exogenous. Since we focus on situations where the exogeneity of instruments is suspect, we limit most of our discussion to 2SLS models where consequences of poor instruments are constrained to the specific equation under examination. In general, researchers estimating simultaneous-equation models should provide 2SLS results in addition to potentially more efficient methods.

As discussed above, 2SLS allows us to separately estimate each of the equations of the simultaneous-equation model. This effectively simplifies the analysis to a single-equation model, and for simplicity our analysis focuses on a single-equation model. This single-equation setting also allows us to examine other econometric issues, such as omitted variables, that can be addressed using instrumental variables.[5] The model describing the impact of the predictor variable (x) on the outcome variable (y) is the following:

$$y = \beta x + \sum_{k=1}^{K} \gamma_k x_k + u \tag{2}$$

As before, x is the explanatory variable of interest (possibly endogenous) and the x_k's are the control variables (assumed to be exogenous). To simplify notation we redefine our x and y variables as being measured *after* controlling for other (exogenous) independent variables.[6] That is, both x and y, are replaced by the residual from a regression of x and y on the x_k's. Besides simplifying the notation, this procedure also highlights that only the unique variation in x and y, after taking into account the control variables, matters in evaluating the regression estimates of interest. We can rewrite the model as follows:

$$y = \beta x + u \tag{3}$$

It is well known that when the correlation between x and u is equal to zero, the OLS estimate of β is consistent (i.e., it asymptotically approaches the true coefficient). This can be seen from the probability limit (plim) of the OLS estimator:

$$\operatorname{plim} b_{OLS} = \beta + \frac{\operatorname{cov}(x,u)}{\operatorname{var}(x)} = \beta + \frac{\sigma_u}{\sigma_x}\operatorname{corr}(x,u) \tag{4}$$

In Eq. (4), var(x) and σ_x are the variance and standard deviation of x, and cov(x, u), and corr(x, u) are the covariance and the correlation between x and u. If x is exogenous (i.e., uncorrelated with u), the second term in Eq. (4) is equal to zero and the OLS estimator is a consistent estimator of the true coefficient. However, this assumption will not be satisfied when there are determinants of x that also affect y directly and these other determinants are not included as control variables in the regression model. When x and u are correlated, the second term in Eq. (4) is not equal to zero and the OLS estimate of β will be inconsistent.

If x and u are correlated, the typical textbook prescription is to use instrumental variables (e.g., Wooldridge, 2002; Greene, 2003). That is, it is necessary to incorporate a variable (z) that is correlated with x but not with u. As with x and y, z is measured after partialling out the effect of the control variables. If such a variable exists, Eq. (3) can be estimated using instrumental variables estimation (IV). The resulting estimator will be consistent, as can be seen from the probability limit of the IV estimator:

$$\operatorname{plim} b_{IV} = \beta + \frac{\operatorname{cov}(z,u)}{\operatorname{cov}(x,z)} = \beta + \frac{\sigma_u}{\sigma_x}\frac{\operatorname{corr}(z,u)}{\operatorname{corr}(x,z)} \tag{5}$$

If the instrument, z, is exogenous (i.e., corr(z, u)=0) and correlated with x, then the second term is zero and the IV estimator is a consistent estimator of the true coefficient. This holds even for small corr(x, z), as long as corr(x, z) *is not zero.*

Two potential problems can arise with the instrumental variable estimator. First, if the instrument is not actually exogenous (i.e., z and u are correlated), the IV estimate moves further away from the true value of β as the correlation

[5] We will not separately discuss dummy-endogenous variables and selection models based on Heckman (1978, 1979). While the general message of our paper also holds for those models, their non-linearity offers special challenges and opportunities. A recent discussion of these models in the accounting literature is provided by Chaney et al. (2008) and Francis and Lennox (2008).

[6] This simplification is without loss of generality, as these two representations result in identical parameter estimates following the Frisch–Waugh theorem (e.g. p. 27 of Greene (2003)).

between z and u becomes larger. This problem is exacerbated when corr(x, z) is small (i.e., when the instruments' first-stage explanatory power is small). We discuss a method to ascertain whether the OLS or the IV estimator is the preferred choice in Section 3.2.

Second, in finite samples, when corr(x, z) is close to zero, the estimates are very unreliable and the standard test statistics are misspecified. This outcome occurs even if the instrument is exogenous (i.e., the expected value of corr(z, u) is zero). The reason is that in finite samples, corr(z, u) is not exactly equal to its expected value. This so-called weak-instrument problem is discussed further in Section 4.

Bartels (1991) also derives the asymptotic mean square error of the instrumental variable parameter estimate as:

$$[\sigma_u^2/n\sigma_x^2][1/R_{xz}^2][1+nR_{zu}^2] \tag{6}$$

where R_{ij}^2 is the squared population correlation between variables i and j, and n is the sample size. The first term in Eq. (6) is the asymptotic mean squared error of the OLS regression and the second term (which is greater than one) is related to the loss in efficiency caused by using an IV estimate as opposed to an OLS estimate. The third term is related to the bias in the IV estimator caused by the use of inappropriate instruments. It is interesting to highlight that when the instrument is actually exogenous, the ratio of 2SLS and OLS standard errors is equal to the correlation (in absolute value) of the instrument z with the original variable x or $|R_{xz}|$. Thus, the power associated with IV estimation can be substantially less than that for OLS.

3.2. Semi-endogenous instruments

Finding a truly exogenous variable that is also correlated with the x is a difficult task for applied researchers because there is typically little theoretical structure to guide this choice. As discussed by Bartels (1991), it is useful to understand whether a "semi-endogenous" variable (i.e., an instrument that is "somewhat" correlated with the error term in the structural equation) will produce IV estimates that are preferred to OLS estimates. We know from Eq. (3) that the resulting IV estimator will not be consistent, but the IV estimate may still have an asymptotic bias that is smaller than the bias in the OLS estimate.

It is possible to identify the circumstances where the bias in the IV estimator is smaller by comparing the bias terms in Eqs. (4) and (5). The IV estimator has smaller absolute bias if the following holds:

$$\frac{\sigma_u}{\sigma_x}\left|\frac{\mathrm{corr}(z,u)}{\mathrm{corr}(x,z)}\right| < \frac{\sigma_u}{\sigma_x}\left|\mathrm{corr}(x,u)\right| \tag{7}$$

Rearranging and simplifying Eq. (7) yields the following condition for the superiority of the IV estimator over the OLS estimator:

$$\left|\mathrm{corr}(z,u)\right| < \left|\mathrm{corr}(x,z)\right|\left|\mathrm{corr}(x,u)\right| \tag{8}$$

We can rewrite this in terms of squared correlations to facilitate comparison with regression output such as the first-stage (partial) R-squared:

$$R_{zu}^2 < R_{xz}^2 R_{xu}^2 \tag{9}$$

As can be seen from Eqs. (5) and (6), the "relative endogeneity" of x and z, and the correlation between x and z are the critical determinants of whether IV estimators are preferred to OLS estimators. For example, if the R^2 in the first-stage regression is 10% (i.e., $R_{xz}^2 = 0.10$), then the correlation between z and u can be no more than 10% of the correlation between x and u for the IV estimation results to be statistically preferred to the OLS results. If the instrument selected by the researcher is moderately to highly correlated with x (which can be tested) and a compelling theoretical or practical argument can be made regarding why the instrument is considerably more exogenous than x, then the IV estimator will be preferred to the OLS estimator. However, if the correlation between the instrument and x is low, then the hurdle for IV estimation to be preferred becomes very high. Even a small correlation between the instrument and the error term will result in large biases in the IV estimator making it more likely that the OLS estimator is preferred (in terms of bias) to the IV estimator.

3.3. Testing the appropriateness of instruments

Several tests have been developed to ascertain the need for and the appropriateness of instrumental variables estimation (e.g., Chapter 6 in Wooldridge, 2002). The most commonly used is the Hausman test which provides a formal test on whether the IV estimator is significantly different from the OLS estimator (Hausman, 1978). Under the assumption of the appropriateness of the instruments, this test can be used to determine the existence of an endogeneity problem, and thus the appropriateness of using OLS. This test statistic can also easily be computed by including both the observed x and the predicted x from the first-stage regression into an OLS version of the second-stage regression. If the coefficient on the predicted x is significant, the Hausman test rejects the null of no endogeneity problem. Variations of this test are applied in the majority of the accounting papers that we have surveyed. If researchers estimate a simultaneous-equation model using

3SLS or maximum likelihood, an additional Hausman test is available for determining whether reliance on potentially more efficient IV applications versus the more robust 2SLS is justified.

Ideally, one would like to have data on the structural error terms, so that the correlation between the instruments and the second-stage error can be measured. Unfortunately, the unobservability of the structural equation error term renders this test impossible. However, it is possible to correlate the instruments with the estimated error term in the second-stage equation. In case of *over-identified* models (i.e., where the number of instruments exceeds the number of endogenous regressors), this test can be used to determine the appropriateness of the instruments under the assumption that at least one of the instruments is *valid* (see also Hausman, 1978).[7] This test should be performed before the Hausman test, as the latter is not valid if the over-identifying restrictions test rejects the appropriateness of the instruments (e.g., Godfrey and Hutton, 1994).

The over-identifying restriction test statistic can be obtained by a regression of the second-stage residuals on *all* exogenous variables. If the instruments are valid, the coefficients on the instruments should be close to zero. The formal test is based on the R^2 from this model being close to zero. In particular, nR^2 is distributed χ^2 with F–G degrees of freedom, where n is the number of observations, F is the number of exogenous variables *unique* to the first-stage, and G is the number of endogenous explanatory variables. It is very important to note that this test requires that at least one of the instruments be valid. If this assumption is violated, there can be a situation where the instruments have similar bias, so that the test will not reject (even in large samples), even though the IV estimates are severely biased.

A limitation of these tests is that their usefulness depends on the power of the test. With low power even invalid instruments may not be rejected, and when the power is very high even unimportant differences will lead to rejection. Therefore, it is desirable to supplement the formal test with a sensitivity analysis that examines whether the use of different instrumental variables yields very different results (we illustrate this latter test in Section 6).

4. Finite sample properties of instrumental variable estimators

4.1. Overview of the weak-instrument literature

Although asymptotic analysis is straightforward to compute and provides important limiting results, it does not provide the necessary insights into the *finite* sample properties of OLS and IV estimators. While the IV estimates are consistent, they are not unbiased (i.e., in finite samples the expected value of the estimate is not equal to the true value). Richardson (1968) and Sawa (1969) provide the exact finite sample properties of a class of IV estimators. They show that the finite sample bias of the IV estimator is in the same direction as the bias in the OLS estimator. Similar to our earlier discussion, this bias is exacerbated when the researcher has weak instruments.

In addition, Nelson and Startz (1990a, 1990b) find that the asymptotic distribution of the IV estimator is a very poor approximation to the finite sample distribution when the instruments are weak. The effect of these two problems is that standard inference based on t-statistics and Wald statistics can be highly misleading in the case of weak instruments (e.g., Bound et al., 1995; Staiger and Stock, 1997; Hahn and Hausman, 2003; and others). Finally, although the econometrics literature often discusses the case where weak instruments lead to over-rejection of the null hypothesis, this is not necessarily the case. Empirical rejection rates can be either much higher or much lower than the pre-specified size of the tests depending on the severity of the endogeneity problem in x (e.g., Hall et al., 1996).

A simple way to detect the presence of weak-instrument problems is to look at the first-stage F-test that the instruments are jointly zero (or partial F-test if there are other control variables). If the F-statistic is low, this implies that the selected instruments are weak. In their survey of the weak-instrument literature Stock et al. (2002) develop some benchmarks for the necessary size of the F-statistic. When the number of instruments is 1, 2, 3, 5, 10, the suggested critical F-values are 8.96, 11.59, 12.83, 15.09, and 20.88, respectively. If the first-stage (partial) F-statistic falls below these critical values, the instruments are considered to be weak and inference problems are potentially serious. The fact that the critical values are increasing in the number of instruments also clearly implies that adding additional low quality instruments is not the solution to a weak-instrument problem. In contrast, the critical F-values usually decrease with the numerator degrees of freedom in tests for model significance.

The econometrics literature has developed several ways of dealing with these problems. Zivot et al. (1998) show that tests based on the Likelihood Ratio (LR) or Lagrange Multiplier (LM) have the correct rejection rates in the single instrument case, whereas the standard test based on the Wald statistic does not. However, in the case of multiple instruments this approach no longer works. Recently, Moreira (2002) demonstrates that one can overcome this problem by adjusting the critical values of the test statistic through simulation of the conditional distribution of the test statistic. The resulting test has correct size even in the presence of weak instruments.

[7] This will be zero by construction for *just-identified* models (i.e., when the number of instruments is equal to the number of endogenous regressors). This observation highlights the importance of having multiple IVs for an empirical application. The key comparison is between the structural error term and the instrument, not between instrument and the dependent variable in the second stage. In our survey of the accounting literature, we found several instances where the authors correlated the instrument(s) with the dependent variable (with or without controls) instead of the error term. Upon finding an insignificant relation, they concluded that they had a valid instrument. This is a completely inappropriate procedure.

4.2. Simulation results using weak and semi-endogenous instruments

It is instructive to use Monte Carlo simulations to provide some insight into finite sample estimates where the instruments vary in term of strength and endogeneity. For the purposes of the simulation we need to put a little more structure on the basic regression model. Following the weak-instrument literature, we use the following model:

$$\begin{aligned} y &= \beta x + u \\ x &= \pi z + \nu \end{aligned} \tag{10}$$

The endogenous variable (y) is assumed to be a linear function of the predictor variable (x) and the random structural equation error (u). The estimate of primary interest is the parameter (β). The predictor variable (x) is assumed to be a function of the instrument (z) through the coefficient π, and other potentially endogenous determinants (ν). As discussed in Section 3, if z is uncorrelated with u, the instrument is strictly exogenous and the IV estimator of β converges to the true value. If both z and ν are uncorrelated with u, x is strictly exogenous and the OLS estimator is consistent.

For this model, the IV estimator can be defined as follows:

$$b_{IV} = \beta + \frac{\hat{\text{cov}}(z,u)}{\hat{\text{cov}}(x,z)} = \beta + \frac{\hat{\sigma}_u}{\hat{\sigma}_x}\frac{\hat{\text{corr}}(z,u)}{\hat{\text{corr}}(x,z)}, \tag{11}$$

To indicate that these are sample covariances not population covariances, we have added the "hats" to the variables in Eq. (11). The second part of the equation is the expression for the bias in the IV estimator, which will asymptotically approach zero under the assumption that z and u are uncorrelated and z and x are correlated, i.e., π is non-zero. However, in finite samples, the estimator is *not* well-behaved because there is a discontinuity in b_{IV} at the point where $\hat{\text{corr}}(x,z)=0$. Depending on the sign of $\hat{\text{corr}}(z,u)$, the bias in b_{IV} approaches positive or negative infinity. Thus, the moments for b_{IV} do not exist.

We assume that z, u, and ν have a normal distribution with a population mean of zero and variances equal to $\sigma_z^2, \sigma_u^2, \sigma_\nu^2$, respectively. For ease of interpretation we will set $\sigma_{z\nu}$ equal to zero. However, we allow both z and ν to be correlated with u. Similar to Section 3.1, one can view all variables as measured after partialling out the effects of any control variables. Thus, even though this structure appears simple, it is sufficiently complex to illustrate the finite sample issues associated with IV estimation.

Rather than directly choosing the parameters, $\sigma_z^2, \sigma_u^2, \sigma_\nu^2, \sigma_{zu}, \sigma_{u\nu}, \sigma_{z\nu}$, and π, we instead pick the desired population correlations, $\rho_{xu}, \rho_{xz}, \rho_{zu}$, and then calculate the parameters necessary to obtain the desired correlations. This allows for a more natural interpretation of the results than would be the case if we directly picked the parameters. We use ρ_{ij} to indicate the population correlations rather than the generic corr(i,j) to highlight that these are population correlations and not the actual correlations in each iteration of the simulation. For each simulation, we use a sample size of 100 observations and set the true β equal to zero. To get the distribution of estimates, we generate 1000 independent samples for each combination of the indicated *population* moments.

For each set of parameters, we present the 5th, 50th, and 95th percentile of the OLS estimator and the IV estimator and the empirical rejection rate of a Wald test at the 5% significance level (tested against the null of β=0). For the case of one regressor, the Wald test is simply the square of the t-statistic and will give equivalent results to the standard t-test (to see this, simplify the general expression for the Wald test, e.g., Wooldridge, 2002, p. 44). Similar results are reported for the 2SLS estimator. As discussed earlier, the standard Wald statistic can have very poor properties when instruments are weak. Hence we also provide the rejection rates based on the Likelihood Ratio (LR) statistic, the conditional Wald statistic and the conditional Likelihood ratio statistic. The latter two are calculated using Moreira's (2002) conditional approach (where the number of iterations used to obtain the conditional test statistics is set equal to 1000).[8] Strictly speaking we do not need to calculate the conditional LR critical value, as it should equal its asymptotic value (3.84) in the case of a single instrument. However, we provide it here to evaluate the adequacy of the number of iterations used to simulate the conditional distribution. If each of these three tests is well specified, the empirical rejection rates should be close to the size of the test, 5%.

The results of the simulation analyses when the *true* β is equal to zero are reported in Table 3. In the first seven rows, we examine the classic weak-instrument case. That is, the variables are drawn from a population where the instrument is strictly exogenous (ρ_{zu}=0), but only weakly correlated with the x-variable (ρ_{xz}=0.01). This latter correlation is set such that while not completely random, the instrument nevertheless is usually insignificant in the first-stage regression and the first-stage F-statistic is below the critical value for a weak instrument (F=8.96 for the case with one instrument) as developed in Stock et al. (2002). Note that while the population moments are indicated as above, in any simulations the actual correlations will vary and can be either positive or negative. It is these finite sample variations that will give rise to the weak-instrument problem.

In these simulations we first vary the endogeneity in x (ρ_{xu}). As should be expected, when ρ_{xu} is equal to zero (i.e., there is no endogeneity problem), the median OLS estimate is equal to zero and the OLS rejection rate of 4.3% is close to the

[8] Moreira and Poi (2001) provide a description of the Stata code needed to perform these tests. This paper and the Stata code can be downloaded at: http://www.columbia.edu/~mm3534/.

Table 3
Simulation results for OLS and IV estimators.

This table shows the results of the endogenous instrument simulation. Each row shows the results based on 1000 independent samples with the indicated *population* moments. We use a sample size of 100 observations and set the true β equal to zero. For each set of parameters we first show the 5th, 50th, and 95th percentile of the OLS estimator, and the empirical rejection rate of a Wald test at the 5% significance level (tested against the null of $\beta=0$). We then do the same for the 2SLS estimator. We also provide the rejection rates based on the Likelihood Ratio (LR) statistic, the conditional Wald statistic (Wald*) and the conditional Likelihood ratio statistic (LR*). The latter two are calculated using Moreira's (2002) conditional approach. All rejection rates are given as a percent. For the purposes of this simulation we set the number of iterations used to obtain the conditional test statistics equal to 1000.

			OLS				**2SLS**						
			Percentiles				*Percentiles*			*Rejection rates*			
ρ_{zu}	ρ_{xz}	ρ_{xu}	0.05	0.50	0.95	Wald	0.05	0.50	0.95	LR	Wald	LR*	Wald*
0.0	0.01	0.0	−0.16	0.00	0.17	4.7	−6.95	0.00	8.26	5.6	0.0	2.0	4.2
0.0	0.01	0.1	−0.06	0.11	0.25	17.0	−6.46	0.10	6.75	5.6	0.0	2.3	3.9
0.0	0.01	0.3	0.14	0.30	0.46	86.9	−5.16	0.26	6.30	4.8	0.2	2.3	2.9
0.0	0.01	0.5	0.35	0.50	0.64	100.0	−4.16	0.45	7.06	4.0	1.0	1.4	2.3
0.0	0.01	0.7	0.58	0.70	0.82	100.0	−4.36	0.70	4.53	5.2	7.9	2.2	1.9
0.0	0.01	0.9	0.83	0.90	0.97	100.0	−1.83	0.89	3.55	5.5	26.9	2.0	1.0
0.0	0.01	0.99	0.97	0.99	1.02	100.0	−0.11	0.98	2.24	5.6	58.0	2.4	1.6
0.0	0.3	0.0	−0.17	0.00	0.16	6.8	−0.57	0.01	0.62	5.2	1.5	3.7	4.8
0.0	0.3	0.1	−0.06	0.09	0.25	15.8	−0.66	−0.01	0.53	4.9	1.2	3.5	4.2
0.0	0.3	0.3	0.12	0.28	0.43	84.9	−0.70	−0.01	0.49	4.3	2.4	3.2	4.0
0.0	0.3	0.5	0.33	0.47	0.61	99.9	−0.90	0.00	0.45	4.6	3.9	3.8	2.9
0.0	0.3	0.7	0.55	0.67	0.79	100.0	−0.99	0.00	0.42	4.8	7.6	4.0	3.7
0.0	0.3	0.9	0.79	0.86	0.93	100.0	−1.04	−0.01	0.35	4.7	7.9	4.7	2.8
0.1	0.1	0.1	−0.05	0.10	0.28	17.7	−5.40	0.65	5.45	17.6	0.8	9.1	17.3
0.1	0.1	0.3	0.14	0.29	0.45	86.6	−3.41	0.76	5.74	18.2	3.7	10.7	17.0
0.1	0.1	0.5	0.35	0.50	0.63	100.0	−2.86	0.76	5.64	16.7	9.0	9.5	12.4
0.1	0.1	0.7	0.58	0.70	0.83	100.0	−2.29	0.87	4.02	19.4	22.9	11.1	12.0
0.1	0.3	0.1	−0.06	0.09	0.26	16.5	−0.23	0.32	1.00	15.1	6.3	12.5	18.2
0.1	0.3	0.3	0.14	0.29	0.43	87.2	−0.29	0.32	0.94	19.4	16.9	16.4	22.8
0.1	0.3	0.5	0.34	0.47	0.61	100.0	−0.28	0.31	0.80	18.0	23.1	15.1	22.0
0.1	0.3	0.7	0.55	0.66	0.77	100.0	−0.25	0.32	0.71	17.2	30.9	14.3	19.2
0.1	0.5	0.1	−0.07	0.08	0.23	18.3	−0.12	0.17	0.48	17.5	14.0	16.4	17.9
0.1	0.5	0.3	0.11	0.26	0.40	85.6	−0.15	0.17	0.45	16.4	17.3	15.4	20.0
0.1	0.5	0.5	0.31	0.44	0.56	100.0	−0.16	0.18	0.42	19.8	24.9	18.9	25.1
0.1	0.5	0.7	0.50	0.61	0.71	100.0	−0.14	0.17	0.39	16.5	26.5	16.1	22.1

nominal rejection rate of 5%. Since the OLS is unbiased in this case, we also find the expected result that the median 2SLS estimate is close to zero. However, due to the low power of 2SLS when the instrument is weak, the range of estimates is very large. Moreover, with weak instruments the standard test statistics become misspecified. The Wald statistic has poor size because none of the 1000 simulations produce a significant 2SLS coefficient. In contrast, the Likelihood Ratio (LR) statistic has a rejection rate close to the 5% level. The two conditional tests perform better than the Wald test, but not quite as well as the LR statistic.[9]

As the correlation between x and u increases, the bias in the median OLS estimate and the rejection rates both increase. This is the standard endogeneity problem. Once the endogeneity problem is large enough, the OLS Wald statistic always rejects the null of zero. As discussed in the weak-instrument literature, 2SLS does not provide a solution in these cases, because the bias in the median 2SLS estimate also increases when the correlation between x and u increases, and is always close to the bias in the OLS estimate. This occurs even though the instruments are perfect (i.e., ρ_{zu} is equal to zero in the population). The reason is that in finite samples, there will always be some spurious correlation between z and u.

For example, consider the case in which the OLS estimate is biased upwards (i.e., x and u are positively correlated). In that setting, when z and u happen to be positively correlated, it becomes more likely that x and z are also positively correlated, since x and u are positively correlated. This results in an upward bias in the 2SLS estimate. Similarly, when z and u happen to be negatively correlated, it becomes more likely that x and z are also negatively correlated, since x and u are positively correlated. This also results in an upward bias in the 2SLS estimate. Therefore, the IV estimates will be biased upwards, just as the OLS estimate.

[9] In the case of one instrument, the conditional LR should be equal to the traditional LR statistic if enough iterations are used to compute the conditional distribution. This result means that 1000 iterations are not quite enough to obtain the correct rejection rates. Given the large number of simulations and the time required to complete each run, we use 1000 iterations per simulation. This is a limitation to our simulation analysis, but we believe that the comparative results for the conditional LR tests are reasonable. For the application tests in Section 6, we will use 100,000 iterations for each test.

The standard Wald statistic also increasingly, and incorrectly, rejects the null hypothesis that β is equal to zero, and for simulations with high endogeneity the test statistic rejects far too often. In contrast, the standard Likelihood Ratio test (LR) continues to perform well, and the conditional Wald and LR statistics also have rejection rates close to the pre-specified 5% level. Thus when the instrument has little explanatory power in the first stage, IV estimation is highly questionable.

For contrast, the second set of estimates still has exogenous instruments, but now the correlation between x and z is much higher at 0.3. This shows the case in which 2SLS works well, the median estimate is very close to the true value and the test statistics have approximately the right size. However, although the 2SLS estimate is consistent in this setting, the estimator is not unbiased in the traditional sense. That is, the expected value of the 2SLS estimate is not equal to zero. Also note that the range of estimates is high, and thus the power of the 2SLS is low. That is, when using 2SLS it is harder to reject the null hypothesis that beta is zero when it is false. Given that in our example the true coefficient is zero, and the null hypothesis is true, this is not harmful. However, in the more general case where the true coefficient is not necessarily equal to zero, the low power of 2SLS is more problematic.

We next illustrate the effects of introducing endogenous instruments into the simulation. In particular, we now draw observations with a relatively small correlation between z and u (ρ_{zu}=0.1). In the three remaining sets of simulations, we vary the strength of the instrumental variable (ρ_{xz}=0.1, 0.3 and 0.5), and within each set we vary the endogeneity in the regressor x (ρ_{xu}=0.1, 0.3, 0.5 and 0.7). In the first set of results where the instrumental variable is relatively weak (ρ_{xz}=0.1), the OLS results are preferred because they have lower bias and smaller dispersion. As strength of the instruments increases and x has moderate to high correlation with u, the 2SLS estimates begin to exhibit lower bias and lower rejection rates than OLS estimates. The dispersion of the 2SLS estimator also drops considerably, alleviating concerns about power. Thus, even when the instrument is semi-endogenous, 2SLS *might* be preferred over OLS when the instruments are reasonably strong. The comparative bias results in these three set of results are consistent with the asymptotic analysis in Eq. (9). For example, the case where 2SLS starts being preferred over OLS is predicted by Eq. (9), since $R^2_{zu} < R^2_{xz} \cdot R^2_{xu}$ in that case (i.e., $0.1^2 < 0.3^2 \cdot 0.5^2$).

The results in Table 3 have implications for the IV methods applied in accounting research. For example, consider the case where the researcher suspects moderate to high endogeneity in the x-variable (e.g., ρ_{xu}=0.3 or 0.5) and the explanatory power for the first-stage regression is low to moderate (i.e., ρ_{xz}=0.1–0.3, implying 1–9% explanatory power). OLS estimates will be preferred to the 2SLS estimates unless the endogeneity in the instrumental variable z is very small (i.e., $\rho_{zu} < 0.1$ or $R^2_{zu} = 0.01$). Although the simulation results are specific to the model in Eq. (10) and selected parameter values, the analysis presented in Table 3 raises serious concern about the desirability of IV estimation methods to correct for endogenous regressor if the selected instruments might be partially endogenous.

4.3. Testing the appropriateness of instruments

Similar to the asymptotic case, it is important to test the appropriateness of the instrumental variables model by performing a Hausman test and a test of over-identifying restrictions. In addition to the issues discussed in Section 3.3, there is also the problem that the small sample distribution of these test statistics can differ quite dramatically from the assumed asymptotic distribution (e.g., Hahn and Hausman, 2003). Absent knowledge about the exact finite sample properties of these tests, simulation can be used to estimate the approximate size of the test statistics for a specific study. For example, Abernethy et al. (2004) perform such a simulation analysis and report that the empirical distribution of the Hausman test in their sample is different from the asymptotic distribution. Finally, it is critical to supplement the formal tests with a sensitivity analysis where the researcher identifies the impact of using different instruments on the IV estimates.

5. Suggested approach for instrumental variable estimation

The analyses in Sections 3 and 4 demonstrate that when the instruments are weak and/or partially endogenous IV methods can produce estimates that are more biased than OLS methods. Moreover, as discussed in Section 2, the typical accounting application does not provide enough information for readers to assess the quality of the IV estimates. Based on the results in Sections 3 and 4, we believe the approach outlined in Table 4 should be used in accounting research.

The first step is to describe the economic theories the research questions is based on. By carefully discussing the economic theories underlying the hypotheses and potential alternative hypotheses, the precise nature of the endogeneity problem will become clearer. For example, a key control variable suggested by one of the theories might be unavailable or hard to measure, the primary regressor could be a choice variable, and/or the direction of causality could be unclear. The more detailed this description, the better the researcher can select an empirical approach to potentially mitigate the endogeneity problem and the better readers can evaluate whether the approach is appropriate.

As part of description of the nature of the endogeneity problem, we recommend that researchers discuss the direction of the bias whenever possible. If the researcher can sign the direction of the bias and the sign is inconsistent with the hypothesis being tested, it may be acceptable to ignore endogeneity because the statistical tests are conservative (although the magnitude of the effect is misestimated). However, frequently the opposite situation will hold. For example, consider a

Table 4
Suggested steps for dealing with endogeneity problems.

Addressing endogeneity problems
• Describe the nature of the endogeneity problem
• Explore alternative research designs
Implementation of the instrumental variable estimation
• Use economic theory to select and justify the choice of instruments
• Evaluate the first-stage results and diagnostics
• Evaluate the second-stage results and diagnostics
• Run a sensitivity analysis on the choice of instruments
• Compare and contrast the estimates from OLS and 2SLS methods
Assess the potential impact of unobserved confounding variables

setting where proprietary costs lead risky firms to disclose lower quality public information and risky firms have a higher cost of capital. Then, ignoring firm risk in the analysis will bias the OLS estimate of the effect of disclosure on cost of capital downwards, favoring the hypothesis that improving disclosure quality leads to lower cost of capital.

The second step is to explore the various alternative ways of solving the econometric problem. Instead of using instrumental variables, researchers can incorporate additional control variables or fixed effects that mitigate the endogeneity problem. Ideally, one would try to use natural experiments, but these are rarely available to the researcher. For example, fixed effects are unlikely to work in the disclosure example because time-series variation in variables such as disclosure quality is often small and related to changes in firm performance. Moreover, prior literature argues that a commitment to disclosure rather than a one-time improvement in disclosure is what drives the cost of capital effect (e.g., Brown et al., 2004).

If instrumental variable estimation is considered to be the most promising econometric approach, the challenge is to find and justify the instruments. For step three, the researcher must use economic theory to argue that the instruments are predicted to affect the *x*-variable, and that they are not correlated with the second-stage error term. A good instrument not only has to be from "outside the system," but it must also not affect the *y*-variable in any way other than through the *x*-variable. While it is always difficult to argue why variables are not correlated, there are some steps authors can take enhance the reader's confidence in the choice of instrument. First, authors should articulate their theory in a manner that explains their selection of the instrument, not merely select the instrument and then attempt to justify this choice. Second, the authors can try to anticipate the potential reasons why the instrument is not exogenous and demonstrate that these effects are either very small or controlled for by inclusion of other variables in the model.

These stringent requirements suggest that several approaches commonly used in the accounting literature, such as industry averages, ranked endogenous regressors, or lagged endogenous regressors are unlikely to be adequate instruments. For example, consider the use of ranked endogenous variables (or portfolio ranks) as an instrument. By construction both the exogenous and the endogenous parts determine the rank (or portfolio assignment). Thus, the instruments obtained this way will have the same problems as the underlying endogenous regressor.

Using industry aggregates as instrumental variables does not generally resolve direction of causality or correct for omitted variables. The reason is that industry aggregates combine both the exogenous and the endogenous parts of the original variables. This approach will only work when the exogenous part of the original variable varies across industry, whereas the endogenous part varies only within industry. However, the more likely scenario is that both parts of the original variable vary across industry. For example, suppose risky, innovative firms provide low quality disclosures to protect their proprietary information. These characteristics are likely related to the industry the firm operates in, and thus other firms in the industry will have the same problems. This endogenous aspect of the variable will not average out when aggregating to the industry level.

Another drawback of using industry averages as an instrument is that it precludes the use of industry fixed effects which is frequently used in the accounting literature to control for unobserved heterogeneity across industries. Implicit in using industry fixed effects is the assumption that the variation *across* industry is more likely to be endogenous than the variation *within* industry, the exact opposite of what is needed for industry to work as an instrument.

Similar to using industry aggregates as the instrument, when using lagged endogenous regressors as the instrument the researcher implicitly makes an assumption about the behavior of the exogenous and the endogenous part of the regressor. When using lagged values of the endogenous regressor, the implicit assumption is that the exogenous part of the regressor persists over time, but that the endogenous part does not persist over time. Using the example above, it is likely that the risk profile and the need to protect proprietary information are relatively stable over time. In that case, using lagged regressors as instruments does not solve the endogeneity problem. The inference from this discussion is not that instruments based on industry averages or lagged regressors never work, but rather that researchers need to clearly delineate why the narrow circumstances in which they do work apply to their research setting.

The ideal instrument is the result of a "natural experiment," event that changes the endogenous regressors, but leaves the other aspects of the economic system unaffected. Obviously, even here a solid understanding of the economic theories is necessary in order to know whether the other aspects of the economic system are really unaffected. In labor economics, probably the area in economics where the use of instrumental variables is most developed, researchers often try to find natural events that influence the variable of interest, such as quarter of birth, rainfall, and the number of rivers and streams running through cities. In accounting, researchers do not typically have access to such variables, but they often use changes in regulation as a quasi experimental treatment (e.g., studies on Regulation FD and Sarbanes-Oxley). One concern about such studies is that it is difficult to disentangle the effect of the specific legislation from all other developments occurring around the same time (including but not limited to the corporate events that drove the new legislation). However, to the extent that the legislation has arbitrary exemptions for, say, firms of certain industries or firms below a certain size, this can be used to strengthen the causal inferences.

To illustrate the difficulty in finding a good instrument even when a natural experiment is available, consider the classic labor economics question regarding the effect of education on lifetime earnings. The primary econometric problem is that education is not randomly assigned to the sample of people because students of high intelligence are more likely to go to college and they also tend to have higher lifetime earnings. In this example, the goal is to find an instrument that predicts education but is not correlated with ability or other parts of the error term. One potential choice is to use variation in quarter of birth which affects schooling through compulsory school attendance laws (e.g., Angrist and Krueger, 1991). Quarter of birth matters because compulsory school attendance laws specify till which age one has to go to school, rather than how many years of schooling one is required to get. The number of (fractional) years of schooling attained at reaching this age limit depends on quarter of birth. This could plausibly be viewed as a random variable and possibly an adequate instrument for the study. However, other studies have linked quarter of birth to various mental problems such as manic depression and schizophrenia, which are likely to affect earnings as well (see Bound et al., 1995; Rosenzweig and Wolpin, 2000). This suggests that even 'natural experiments' can fall short of the standard necessary to employ instrumental variables.

Only after the problem has been clearly defined and the instruments have been carefully selected, does one move to the actual estimation. This starts with the estimation of the first stage. This is not just a necessary step in the 2SLS procedure, but it also provides vital diagnostic information. In this first stage, all independent variables should be included, not just the selected instruments. These first-stage results should be reported and researchers should discuss whether the instruments have the expected signs, magnitudes that are reasonable, and coefficients that are statistically significant at conventional levels.

In addition there are a number of diagnostic tests that should be reported. To assess whether weak instruments are a problem, researchers should report partial F-statistic and partial R^2. If the instruments are weak, it is necessary to use the results in Moreira (2002) for significance tests. The first-stage R^2 also sets the upper bound for the acceptable degree of endogeneity in the instrumental variables (i.e., where 2SLS is preferred over OLS in terms of bias) as illustrated by Eq. (9). Once the correlation between the regressor and the instruments is computed, it is necessary for the researcher to explain why the endogeneity of the instruments is substantially below the endogeneity of the regressor. If the explanatory power of the instruments is low (i.e., the instruments are weak), the researcher should provide justification for whether the instruments are sufficiently exogenous.[10]

If the first-stage results are considered adequate, the second-stage results can be estimated. As part of the second stage, it is desirable to discuss the magnitude of the coefficient on the instrumented variable. While there is some tendency in the accounting literature to simply focus on the sign and statistical significance of the coefficients, the magnitude of the effect is equally important. In particular when the instruments have low explanatory power in the first stage, it is common that the estimated coefficients on the instrumented variable will become unreasonably large or small in the second stage. If this is the case, this would provide clear evidence that the IV estimates are not reliable enough to replace the OLS estimates.

If multiple instruments are available (which is the preferred case), the over-identifying restrictions test should be computed. It is also informative to demonstrate that different instruments (or different sets of instruments) provide the same substantive results. If this is not case, the researcher should question the validity of the instruments. If different instruments or sets of instruments produce very different results, this suggests that some, or all, of the instruments are not exogenous.

The researcher should also report the OLS results so that the reader can compare between IV and OLS estimations. It is important to see whether the coefficient on the endogenous variable changes in the expected direction. For example, if the concern was an omitted variable that would bias the OLS upwards, then the IV estimate should be lower than the OLS estimate. Differences between the OLS and IV estimates can be statistically assessed using a Hausman test.

Finally, the discussion so far has focused on whether the 2SLS results are preferred over the OLS results. However, it may be possible to gain some insights when OLS and 2SLS estimates have the same sign and similar magnitudes and significance levels. Although obtaining similar results with different estimation methods seems desirable, there are two important caveats associated with this type of conclusion. First, the finite sample and weak-instrument literature shows

[10] It is also important to realize that weak instruments are not just a small sample problem. In particular, Bound et al. (1995) find evidence that weak instruments are a problem in the Angrist and Krueger (1991) study despite having over 300,000 observations.

that the IV estimator is biased in the same direction of the OLS estimator even in the case of perfectly exogenous instruments. Second, when the regressor is correlated with the error term, it will generally be the case, that, on average, its determinants are also correlated with the error term in the same direction. This implies that if a random subset of the determinants is used as instruments, the 2SLS estimates will again be biased in the same direction as the OLS estimates. Thus, if the analyses discussed in this paper suggest that the instrumental variables are weak and/or semi-endogenous, the IV estimator should neither be used to replace the OLS estimator, nor should it be used to bolster confidence in the OLS estimate, even if the two results happen to be close.

The above discussion provides a somewhat pessimistic assessment regarding the ability of researchers to identify adequate instrumental variables. If a solution to the endogeneity problem proves elusive, an alternative approach is to assess how large the endogeneity problem must be in order to overturn the results generated using OLS (i.e., without any control for endogeneity). In this approach, there is no attempt to identify instrumental variables because the researcher is simply attempting to assess the sensitivity of the results to unobserved factors (e.g., Frank, 2000; Rosenbaum, 2002; DiPrete and Gangl, 2004, and others). We apply and discuss this approach in more detail in Section 7. If the results are statistically sensitive to small, plausible levels of endogeneity, the researcher needs to be very cautious interpreting the test outcomes. This approach is most fruitful when the nature of the endogeneity problem is caused by unobserved correlated omitted variables. While this approach does not correct the bias induced by endogeneity, it does provide the researcher with some indication about the severity of this econometric problem.

There is no fool-proof way of dealing with the problem of endogeneity in empirical accounting research. However, following the process outlined in this section should help researchers assess whether the use of instrumental variables is likely to mitigate or exacerbate the undesirable impact of endogeneity on the results reported for a specific research question. We illustrate some of these issues in the next section.

6. Corporate disclosure and the cost of capital

In this section, we examine the effect of voluntary disclosure on the firm's cost of capital. This topic has received considerable interest from accounting researchers, but remains controversial from both a theoretical and econometric perspective. While earlier research in this area has treated disclosure as exogenous (e.g., Botosan, 1997), recent work explicitly recognizes the endogenous nature of the disclosure decision. We rely on several papers that have attempted to address the potential endogeneity of the disclosure choice using instrumental variable methods.

The IV applications and results on the determinants of cost of capital vary across research studies. For example, Leuz and Verrecchia (2000), Hail (2002), and Brown and Hillegeist (2007) find the expected negative relation between their disclosure proxies and selected measures for the cost of capital (or information asymmetry) after using IV estimation. In contrast, Cohen (2008) finds that the relation between his measures of reporting quality and cost of capital is no longer significant after taking into account the endogeneity of the choice of reporting quality. The purpose of our analysis is not to resolve this issue, but rather to demonstrate the choices and problematic issues that confront researchers using IV estimation. While our discussion focuses on the endogeneity problem in this setting, other concerns might be relevant as well. For example, measurement error problems for both cost of capital and firm disclosure quality are likely to be substantial and might be as severe as the endogeneity problem we discuss.

Our measure for the cost of capital is developed using several measures of implied cost of capital. Rather than picking one specific metric, we use the average of the four implied cost of capital measures investigated by Guay et al. (2003). They study measures based on Gebhardt et al. (2001), Claus and Thomas (2001), the Gordon growth model, and Gode and Mohanram (2003). We use this index for cost of capital because it is expected to exhibit lower measurement error than any of the four individual measures.

Our disclosure measure is based on the AIMR disclosure ratings. As in Lundholm and Myers (2002), we transform the original scores by ranking firms within industry and subtracting one before dividing by the total number of firms being ranked in the industry for the year minus one. The resulting scores range from zero to one. This implies that each year, the most forthcoming firm in each industry receives a rating of one, and the least forthcoming firm in each industry receives a rating of zero. This transformation allows easy interpretation of the coefficient on disclosure quality, as it is the difference in cost of capital between the most forthcoming and the least forthcoming firm in the industry.

For this setting, endogeneity affects several aspects of the research question. First, there is little consensus in the theoretical literature as to whether disclosure quality is diversifiable or whether it should affect cost of capital (e.g., Lambert et al., 2007). Moreover, there is also no consensus on whether the impact of disclosure occurs through parameter uncertainty or information asymmetry. The absence of a consensus in the finance literature on which risk factors are priced makes it difficult to identify the control variables or select acceptable instruments. In addition, voluntary disclosure is a choice variable that is affected by many factors including any perceived cost of capital benefits.

As a result, the direction of these effects induced by endogeneity is unclear. For example, as discussed earlier, risky innovative firms may have low disclosure quality because their financials are inherently hard to predict and they might want to protect their proprietary information. This would cause a negative relation between disclosure quality and cost of capital (assuming the underlying risk is priced, but not controlled for). In contrast, firms with high risk and uncertainty in their business environment (and thus a high cost of capital) may try to increase their disclosure quality in order to reduce

cost of capital. To the extent that they are only partially successful, this causes a positive relation between disclosure quality and cost of capital. Given the nebulous nature of the endogeneity problem in this case, it is very difficult to assess the direction of the bias induced by endogeneity in the disclosure variable.

In contrast to our recommendations in Table 4, we do not select instruments based on economic theory, but instead simply select the set of variables previously used as instruments for disclosure quality. To our knowledge, a complete structural model based on the costs and benefits associated with disclosure choice has not been developed, and thus it is very difficult to select appropriate instruments in disclosure research. Although we do not use economic theory to justify our choice of instruments, our choice of instruments does enable us to demonstrate the problematic features in the existing literature. The absence of a rigorous economic model to justify the selection of instruments is a major limitation to most accounting research using this methodology and is not restricted to disclosure research.

The instrumental variables obtained from the prior literature are the natural logarithm of the number of owners (defined as number of common shareholders), one-year sales growth, capital intensity (defined as the ratio of property, plant, and equipment to total assets), operating margin (defined as sales minus cost of goods sold, scaled by sales), length of operating cycle (defined as average receivables divided by sales plus average inventory divided by cost of goods sold), and the presence of a Big-six auditor. In addition we use the following control variables: log of market value of equity, book-to-market equity, number of analysts, leverage (defined as total debt/market value of equity), and return on assets (defined as operating income after depreciation scaled by total assets).

Given that our instruments are generic firm characteristics, it is questionable whether they meet the stringent requirements for instruments. For example, Big-Six auditors conceivably could improve the quality of the firms' earnings reports and other disclosure. However, the choice of auditor is endogenous, and it is likely driven by the firm's overall disclosure policy, just as the variable it is supposed to instrument. Similarly, longer operating cycles conceivably make it harder to make accurate predictions of the firm's performance, thus lowering the disclosure quality. However, the length of the operating cycle is also likely related to the risk of firm's operations, thus potentially leading to a downward bias (more negative) in the second-stage coefficient on disclosure quality.

The cost of capital measures are computed as of July 1 for each year from 1982 to 1996. We match this measure with the independent variables calculated using data from the prior fiscal year. In order to reduce the influence of outliers, we truncate all variables at the 1st and 99th percentile. The descriptive statistics are reported in Table 5. The regressions are estimated in a pooled time-series cross-section setting after subtracting the industry-year specific mean from each variable in order to remove temporal and industry effects.

The OLS results obtained from the regression of the cost of capital on disclosure quality and control variables are presented in Table 6. We find that moving from the least forthcoming to the most forthcoming firm yields a statistically insignificant 0.06% decrease in cost of capital. The insignificant coefficient is consistent with findings by Botosan and

Table 5
Descriptive statistics for the variables used in the analysis of the relation between corporate disclosure quality and the cost of capital.
This table shows the descriptive statistics of the disclosure example. The sample is based 2870 firm-year observations from the period 1982 to 1996.

Variable	Mean	Std	Min	Max
Cost of capital	11.60	3.18	5.22	34.27
Disclosure quality	0.51	0.31	0.00	1.00
Log (# owners)	2.57	1.35	−1.98	6.44
Sales growth	0.10	0.16	−0.41	1.52
Capital intensity	0.37	0.23	0.00	0.90
Operating margin	0.36	0.16	0.06	0.92
Operating cycle	1.01	1.80	0.05	9.75
Big-six auditor	0.86	0.35	0.00	1.00
Log market value	7.27	1.33	3.58	11.44
Book-to-market	0.62	0.34	0.06	2.67
# Analysts	9.72	5.45	1.00	30.00
Leverage	0.22	0.12	0.00	0.75
Return on assets	0.11	0.07	−0.08	0.35

Variable definitions: Cost of capital is defined as the average of the following four implied cost of capital measures: Gebhardt et al. (2001), Claus and Thomas (2001), the Gordon growth model, and Gode and Mohanram (2003). Disclosure quality is defined as the AIMR disclosure ratings. As in Lundholm and Myers (2002), we transform the original scores by ranking firms by industry and subtracting one before dividing by the total number of firms being ranked in the industry for the year minus one. The resulting scores range from zero to one. Higher scores imply better disclosure. Log of # of owners is defined as the natural log of Compustat data item #100: Common shareholders. One-year sales growth is defined as sales minus lagged sales divided by lagged sales. Capital intensity is defined as the ratio of property, plant, and equipment to total assets. Operating margin is defined as sales minus cost of good sold, scaled by sales. Length of operating cycle is defined as average receivables divided by sales plus average inventory divided by cost of goods sold. Big-six auditor is a dummy variable that has a value of one if the firm has a Big-six auditor. Log of market value of equity is the natural log of fiscal year-end price times the number of shares outstanding, book-to-market equity is defined as book value of equity (Compustat data item #60 divided by market value of equity. Number of analysts is equal to the number of analysts that issued a forecast for next year's earnings. Leverage is defined as total debt/market value of equity. Return on assets is defined as operating income after depreciation scaled by total assets.

Table 6
The relation between corporate disclosure quality and the cost of capital.

This table shows results from the disclosure study. The first set of results is from a standard OLS regression of cost of capital on disclosure quality and control variables. The second set of results is from a 2SLS regression where disclosure quality is treated as endogenous. The third set of results is from an alternative second-stage regression in which all independent variables have been replaced by the product of the first-stage regression coefficient and the original variable. Unlike the standard 2SLS the coefficients on the instruments are not constrained to be the same. The lower part of the table shows the partial *F*-statistic and the partial *R*-squared from the first-stage regression and the values for the two specification tests: the test of over-identifying restrictions and the Hausman test. And finally, the Wald statistic for the robust test on the 2SLS coefficient on disclosure quality, and the 95% and 99% critical values for the Wald statistic (based on Moreira's (2002) conditional approach using 100,000 simulations).

	OLS		**2SLS**				**Unconstrained**	
			First-stage		*Second-stage*		*Second-stage*	
	Coef	*t*-Stat	Coef	*t*-Stat	Coef	*t*-Stat	Coef	*t*-Stat
Disclosure quality	−0.06	−0.54			−5.31	−2.52		
Instruments								
Log (# owners)			0.00	0.18			183	5.17
Sales growth			−0.06	−1.43			1.97	0.44
Capital intensity			0.05	1.06			−20.38	−3.40
Operating margin			0.07	1.32			−22.06	−4.60
Operating cycle			−0.03	−2.00			−10.72	−3.38
Big-six auditor			0.10	2.40			1.14	0.43
Control variables								
Log market value	0.10	1.75	0.01	0.77	0.15	1.94	−15.58	−1.89
Book-to-market	2.62	14.35	−0.07	−2.27	2.25	8.11	−32.24	−11.55
# Analysts	−0.03	−2.12	0.02	9.34	0.07	1.67	−1.53	−2.25
Leverage	2.46	6.52	0.19	3.22	3.49	5.46	13.12	6.67
Return on assets	−5.14	−5.82	0.05	0.32	−4.61	−3.97	−65.95	−3.38
Adjusted *R*-squared	0.16		0.09		0.11		0.18	
Partial F-statistic			F_p=2.59 (p=0.017)					
Partial R-squared			R^2=0.005					
Over-identifying restrictions test			χ^2=38.01 ($p<0.001$)					
Hausman test			F=10.47 (p=0.001)					
Wald test on Disclosure rating			W=6.45					
Conditional Wald 95% (99%) critical value			W_{crit}=3.06 (W_{crit}=4.67)					

Plumlee (2002).[11] We next estimate a 2SLS regression using the previously discussed instruments. Consistent with standard IV approach, we include *all* exogenous variables in the first-stage. The R^2 of this first-stage model is 9%. However, this overstates the true explanatory power of the instruments as the control variables also contribute to this R^2. After removing the contribution of the control variables, the *partial* R^2 is approximately 0.5% and the partial *F*-statistic of the first-stage model is 2.59. Based on the analysis by Stock et al. (2002) such a low *F*-statistic is indicative of weak instruments and this necessitates the use of robust inference procedures. Therefore we use the conditional test statistics based on Moreira (2002) in addition to the standard test statistics. Apart from causing inference problems, such a low partial R^2 also suggests that small variations in sampling can have large effects on the second-stage results causing large jumps in coefficients between samples and years.

At this point, it is necessary to evaluate whether IV estimation is likely to represent an improvement over OLS. From Eq. (6), we know that the squared correlation of the instruments with the structural error term has to be less than 0.5% (the partial R^2) of the comparable squared correlation between disclosure quality and the structural error for 2SLS to provide better estimates than OLS. Thus, the selected instruments *must* be substantially more exogenous than disclosure quality for the 2SLS estimates to dominate the OLS. In particular, it is necessary that the squared correlation between disclosure quality and the structural equation error is approximately 200 (or 1/0.005) times larger than the comparable squared correlation of the instruments with the structural error term. Although these correlations are unobservable and the inequality cannot be directly tested, we believe that this criterion is unlikely to be met in this particular research setting. In that case, instrumental variable estimates are likely to be more biased than the OLS estimates and should not be used to replace the OLS estimates.

While the above results might lead one to abandon the IV estimation, for the purposes of this illustration, we complete the 2SLS approach by estimating the second-stage. The 2SLS coefficient estimate for disclosure quality indicates that

[11] There is a potential problem with this specification. Since changes in the cost of capital also affect the market valuation of the firm, it is potentially problematic to include contemporaneous market based measures such as the market value of equity and the market-to-book ratio in the regression. In untabulated robustness tests, we consider two alternatives: lagging the market based measures by a year, and using log of total assets instead of the market based measures. The OLS results (and subsequent 2SLS results) are not materially affected by these changes.

moving from the least forthcoming to the most forthcoming firm yields a statistically significant 5.31% decrease in cost of capital. One problem with this test is that the significance level is based on standard asymptotic inference which can be very misleading when the instruments are weak. However, using the conditional approach in Moreira (2002), we find that conditional Wald statistic for the 2SLS coefficient exceeds the simulated 95% and 99% critical values. Thus, the rejection of the null hypothesis is not due to incorrect size of the test caused by weak instruments.

To assess the quality of this 2SLS estimation, it is instructive to assess the size of the coefficient estimate because unrealistically high or low estimates should cause the researcher to be suspicious about the quality of the instruments. The coefficient estimate on disclosure quality (-5.31) is approximately 45.7% of the mean for cost of capital. It seems unlikely that among large firms followed by analysts, an exogenous shift in disclosure quality from the least forthcoming to the most forthcoming in the same industry could have this large effect on the cost of capital.

The common way to justify the use of 2SLS rather than OLS results is to perform the standard Hausman test. For our sample, the Hausman test strongly rejects the exogeneity of disclosure quality (F=10.47, $p<0.001$). Based on this result, accounting researchers typically conclude that the 2SLS estimate is preferable to the OLS estimate. However, the validity of this conclusion critically depends on the appropriateness of the instruments (i.e., whether the instrumental variables are actually exogenous). Thus, the Hausman test alone is *not* sufficient to conclude that the coefficient on disclosure quality is -5.31 and statistically significant at conventional levels as opposed to being -0.06 and not statistically significant.

Since we have multiple instruments for the endogenous variable, we compute a test of over-identifying restrictions. If this test rejects the appropriateness of the instruments, it is not appropriate to proceed to the Hausman test (e.g., Godfrey and Hutton, 1994). Equivalently, one can examine the sensitivity of second-stage estimates to the use of different (sets of) instruments. The intuition of this test is that if the instruments are valid, each instrument should provide similar estimates for the true coefficient.

Our sensitivity analysis is implemented in the last two columns in Table 6, i.e., "unconstrained" second-stage). The model is an OLS regression of cost of capital on all of the independent variables. For ease of comparison, each independent variable is replaced by the *product* of its original value and its associated first-stage coefficient. This facilitates the interpretation because the coefficient on each instrument is equal to the second-stage coefficient on disclosure quality in a model where that instrument is the only instrument and the rest of the instruments are treated as control variables (i.e., they are included explicitly in the second-stage). If the instruments are valid, the resulting coefficients for the instruments should be close to each other, and therefore close to the 2SLS estimate (which is the weighted average of these estimates).

The results in Table 6 illustrate that the coefficients on the assumed exogenous variables vary considerably. For example, if number of owners had been used as the sole instrument, we would have found a theoretically unexpected positive and statistically significant effect of 183% (i.e., the most forthcoming firm would have extremely high cost of capital). However, if capital intensity or operating margin had been used as instruments, we would have obtained an implausibly large estimate of a 20% decrease in cost of capital moving from the most forthcoming to the least forthcoming firm in the industry. As might be expected given the range of estimates produced by the unconstrained approach, the formal test for over-identifying restrictions rejects the exogeneity of the instruments (χ^2=38.0, p=0.001). The sensitivity analysis and the formal over-identifying test indicate that our instruments are of questionable quality, and thus unlikely to produce better estimates and inference than OLS.[12]

There are three issues that should be highlighted with the use of the over-identifying restrictions test. First, Hahn and Hausman (2003) show that the size of the test in finite samples can differ significantly from the asymptotic size, leading to false rejections. Second, in very large samples, this test may be so powerful that economically small deviations lead to rejections (even though 2SLS is much better than OLS). However, in small samples, the test may lack power to reject even economically important deviations (even though 2SLS might well be worse than OLS). Thus, it is important to supplement the formal test with a sensitivity analysis such as our "unconstrained" second-stage to assess the similarities (or dissimilarities) in the coefficient estimates obtained when using different sets of variables as instruments. Finally, neither this test nor the sensitivity analysis will detect econometric concerns when all instruments exhibit similar problems. That is, if the instruments all lead to a bias in the same direction with comparable magnitude, the over-identifying restriction test will not reject the null of valid instruments (even in large samples), but the estimated coefficient for the endogenous variable can be severely biased. Therefore, the over-identifying restriction test should be used as a check on instruments justified by economic theory and should *not* be used to select instruments, unless the researcher has a proper instrument to use as a comparison benchmark.

In the disclosure example we find that the selected instrumental variables provide wildly varying estimates and do not satisfy the over-identifying restrictions test. This result, along with the very low likelihood that the squared correlation of the instruments with the structural error term is approximately 200 times smaller than the comparable squared correlation between disclosure quality and the structural equation error, lead us to conclude that OLS results provide a more appropriate basis for inferential conclusions than the 2SLS results. This holds despite the likely endogenous nature of

[12] An alternative sensitivity test is to estimate the 2SLS model using one instrument at a time, but eliminating the other instruments from the model (rather than using them as control variables). When we use this approach for each of the six instruments, we find that cost of capital estimates range from -28% to $+71\%$ depending on which instrument is used. This supports our argument that the instruments are of questionable quality in this example.

disclosure quality. Thus, in our example, we find no statistical evidence that disclosure quality has any association with the cost of capital. However, the key point of this example is not this result (which might be specific to this sample or a result of poor measures), but rather to illustrate how to assess the quality of the instrumental variable estimation in a typical accounting setting.

7. An illustration of an alternative approach

The prior section showed the some of the difficulties in obtaining reliable instrumental variable estimates. However, when the *x*-variable is endogenous, the OLS estimates are biased as well. In settings where valid instruments variables are not available, the question arises how to evaluate these OLS estimates. As discussed in Section 5, an alternative approach is to assess how large the endogeneity problem has to be in order to change the OLS results. In particular, how large does the problem have to be to make the coefficient statistically insignificant? Unfortunately, in our disclosure example, the coefficient on disclosure quality was already not statistically significant in the OLS regression. Therefore, we extend the example by using effective bid-ask spreads instead of implied cost of capital as our dependent variable. The effect of disclosure on liquidity has a stronger theoretical foundation and has more robust empirical support.

We obtain the trades and quotes from the ISSM and TAQ databases to calculate relative bid-ask spreads. The relative effective spread is measured as the difference between the trade price and the midpoint of the bid price and the ask price, scaled by the trade price. We follow the approach in Ng et al. (2008) to clean and aggregate the transactions data. We then match this data with our sample which results in 1398 firm-year observations for which we have all required data. The results of this analysis are reported in Table 7. Consistent with the theoretical predictions we find a statistically significant negative association between disclosure quality and effective spreads, although the magnitude of the effect is very modest.

Since the estimated OLS coefficient on disclosure quality is statistically significant, we examine the potential impact of *unobserved* confounding variables using the approach in Frank (2000). This method is based on the notion that for an unobserved variable to affect the results it needs to be correlated with both the x-variable and the y-variable (controlling for the other variables). Frank (2000) derives the minimum correlations necessary to turn a statistically significant result into a borderline insignificant result. The Impact Threshold for a Confounding Variable (denoted as ITCV) is defined as the lowest product of the partial correlation between y and the confounding variable and the partial correlation between x and the confounding variable that makes the coefficient statistically insignificant. If the ITCV is high (low), the OLS results are robust (not robust) to omitted variable concerns.

We calculate ITCV for disclosure quality in column (3). The threshold value for disclosure quality is -0.028, implying that the correlations between x and y with the unobserved confounding variable each only need to be about 0.167 ($=\sqrt{0.028}$) for the OLS result to be overturned. Since disclosure is negatively related to spreads, one of these two correlations needs to be negative or else the confounding variable would strengthen rather than weaken the effect.

Without some additional analysis, it is difficult to determine whether the ITCV is small enough to conclude that the OLS association between disclosure quality and spreads is fragile. That is, we need to develop a benchmark for the size of likely correlations involving the unobserved confounding variable. While, by definition, we do not have the unobservable confounding variable, we do have other control variables. In column (4), we show the impact of the inclusion of each independent variable on the coefficient of disclosure quality. Similar to the ITCV the impact is defined as the product of

Table 7
Analysis of the impact of unobservable confounding variables.
This table shows the effects of disclosure quality on the effective bid-ask spread, with an assessment of the impact of unobservable confounding variables based on Frank (2000). Panel A shows the results of a standard OLS regression of effective bid-ask spread on disclosure quality and control variables. The number of firm-year observations is 1398. For each of the independent variable an impact statistic is calculated (ITCV) indicating the minimum impact of a confounding variable that would be needed to render the coefficient statistically insignificant. The ITCV is defined as the product of the correlation between the *x*-variable and the confounding variable and the correlation between the *y*-variable and the confounding variable. To assess the likelihood that such a variable exists, column (4) shows the impact of the inclusion of each independent variable on the coefficient on disclosure quality. The impact is defined as the product of the partial correlation between the *x*-variable and the control variable and the correlation between the *y*-variable and the control variable (partialling out the effect of the other control variables). The sign of the impact measure indicates how the inclusion of the column variable affects the coefficient of the row variable. A positive impact score indicates that inclusion of the column variable makes the coefficient on the row variable more positive (less negative), a negative impact score has the opposite effect. Column (5) includes a more conservative measure of impact, the product of the simple correlation between the *x*-variable and the control variable and the simple correlation between the y-variable and the controlling variable.

OLS regression of bid-ask spreads on disclosure and controls, with impact statistics	(1) Coefficient	(2) *t*-Statistic	(3) ITCV	(4) Impact	(5) Impact_{Raw}
Disclosure quality	−0.09	−2.91	−0.028		
Log market value	−0.29	−19.27		−0.009	−0.110
Book-to-market	0.00	0.00		0.000	−0.036
# Analysts	0.01	2.56		0.007	−0.085
Leverage	−0.04	−0.38		−0.001	0.001
Return on assets	0.21	0.88		0.000	−0.014

the partial correlation between the x-variable and the control variable and the correlation between the y-variable and the control variable (partialling out the effect of the other control variables). The sign of the impact measure indicates how the inclusion of the control variable affects the coefficient of disclosure quality. A positive impact score indicates that inclusion of the control variable makes the coefficient on the disclosure quality more negative (less positive) and a negative impact score has the opposite effect.

The variable with the largest impact on the coefficient for disclosure quality is market value of equity (MVE), with a value of -0.009. This suggests that we would need a confounding variable with a stronger impact than MVE to overturn the results on disclosure quality. Specifically, the unobserved confounding variable must be more highly correlated with disclosure quality and bid-ask spreads than MVE. Under the assumption that we have a good set of control variables this provides some confidence in the estimate of the effect of disclosure quality on bid-ask spreads.

The impact of each control variable is measured after inclusion of the other control variables. As such, even though the correlations between MVE and disclosure quality and MVE and bid-ask spreads are high, the partial correlations are relatively low. In comparing the ITCV to the distribution of impact scores for the control variables we implicitly assume that the confounding variable is similarly correlated with the other control variables. To the extent that the confounding variable is relatively distinct, a more fitting comparison might be to look at the product of the *raw* correlations instead of the partial correlation. Column (5) includes this more conservative measure of impact. Comparing these impact scores to the ITCV suggests that the effect of disclosure quality is not nearly as robust as previously implied. However, one might still argue that even though a variable with similar impact as MVE would overturn the results, it is unlikely such a variable will be found given that we already have strong controls such as MVE and the number of analysts in the model.

The assessment of confounding variables is a very useful evaluation procedure for the OLS estimates. However, absolute standards for impact threshold are difficult to establish.[13] Researchers can make use of the impact for the selected control variables to provide a reasonable benchmark to the ITCV.

8. Concluding remarks

There is little doubt that endogeneity causes substantial econometric problems in virtually all non-experimental empirical accounting research. Accounting researchers are aware of these econometric problems and they commonly use IV (instrumental variable) methods in the hope of mitigating the inconsistency in parameters estimates. However, as we have shown in our synthesis and extension of the contemporary econometrics and statistics literatures, many IV applications in accounting are likely to produce highly misleading parameter estimates and inferential tests. We agree with the insightful comments by Hamermesh (2000, p. 371) regarding IV estimators:

> … its proponents are often too quick to assume that the chosen instrument is exogenous and generates a consistent estimate of the population parameter.
>
> One must be able to argue that the instrument is beyond the decision-makers' control and that it describes behavior that is randomly distributed in the population one wishes to describe.

Assuming that IV approaches are used, there are several fundamental requirements that must be met for any IV application in accounting research. First, it is necessary to describe the nature of the endogeneity problem inherent in the research question and to evaluate various alternative empirical approaches. Second, the choice of instrumental variables must be justified. In particular, the correlation between the instruments and the structural error has a critical impact on the usefulness of the IV estimators and researchers must use economic theory, prior empirical results, and intuition to convince the reader that the size of this correlation is small enough that IV estimators are superior to OLS estimators. Third, it is essential to report the full results of the first-stage regression, including the partial F-statistic and partial R^2. Fourth, it is necessary to compare and contrast the estimates from OLS and 2SLS methods to see whether the differences between the methods are consistent with the underlying theory. Fifth, researchers should provide sensitivity analyses on the instruments and report tests for the appropriateness of the IV approach. Finally, it is important to assess the potential impact of unobserved confounding variables using techniques such as those develop by Frank (2000), Rosenbaum (2002), and DiPrete and Gangl (2004).

Although this paper has mostly considered instrumental variable methods, there are other ways to assess and mitigate the econometric problems caused by endogeneity. A more direct method of coping with endogenous regressors is to develop and estimate a structural econometric model describing the research question.[14] Some researchers consider the choice of instrumental variables to be a purely statistical exercise with little real economic foundation (i.e., find an exogenous variable that is correlated with the endogenous regressor, but at the same time uncorrelated with the disturbance). In contrast, structural modeling consists of specifying the economic model that is assumed to have given rise

[13] It would be useful for future research to develop reference distributions for the typical effect sizes (e.g., partial correlations for the primary variables of interest) in empirical accounting research (see Cohen, 1988, for an example of this type of study in psychology). These reference distributions could be used to assess the likelihood that an observed ITCV is small (large) enough that endogeneity will (will not) affect the estimated OLS coefficient.

[14] A comprehensive discussion of structural econometric models is contained in the excellent survey of Reiss and Wolak (2006) which examines structural econometric modeling in the context of the economics of industrial organizations.

to the observed data (i.e., the data generating mechanism). This requires the researcher to be precise about which variables are exogenous and endogenous, the objective functions and optimization of the choice variables of the economic agents, and the nature of the equilibrium. Once the economic system is modeled, statistical assumptions are then placed on the model which allow for the estimation of the structural parameters of the model (typically using generalized method of moments or maximum likelihood estimation). An important benefit of a structural model relative to an instrumental variables approach is that it enables the researcher to make explicit causal statements with regard to the endogenous variables in the system rather than merely identifying them.[15]

The ultimate question is whether IV estimation is useful for typical empirical accounting research. While it is impossible to answer this question completely, our perspective is that accounting researchers need to be much more rigorous in selecting and justifying their instrumental variables. Moreover, researchers should report diagnostic tests for the first- and second-stage results, along with an assessment for the potential impact of unobserved confounding variables. We hope that our paper provides the necessary background for accounting researchers to collectively decide on the appropriate method for dealing with now the standard "endogeneity critique" of empirical research.

References

Abernethy, M.A., Bouwens, J., Van Lent, L., 2004. Determinants of control system design in divisionalized firms. The Accounting Review 79, 545–570.

Aboody, D., 1996. Market valuation of employee stock options. Journal of Accounting and Economics 22, 357–391.

Aboody, D., Barth, M.E., Kasznik, R., 2004. SFAS No. 123 stock-based compensation expense and equity market values. The Accounting Review 79, 251–275.

Anderson, R.C., Mansi, S.A., Reeb, D.M., 2004. Board characteristics, accounting report integrity, and the cost of debt. Journal of Accounting and Economics 37, 315–342.

Angrist, J.D., Krueger, A.B., 1991. Does compulsory school attendance affect schooling and earnings? Quarterly Journal of Economics 106, 979–1014.

Ball, R., Shivakumar, L., 2005. Earnings quality in UK private firms: comparative loss recognition timeliness. Journal of Accounting and Economics 39, 83–128.

Bartels, L.M., 1991. Instrumental and 'quasi-instrumental' variables. American Journal of Political Science 35, 777–800.

Barth, M.E., Kasznik, R., McNichols, M.F., 2001. Analyst coverage and intangible assets. Journal of Accounting Research 39, 1–34.

Barton, J., 2001. Does the use of financial derivatives affect earnings management decisions? The Accounting Review 76, 1–26.

Barton, J., Waymire, G., 2004. Investor protection under unregulated financial reporting. Journal of Accounting and Economics 38, 65–116.

Beatty, A., Chamberlain, S.L., Magliolo, J., 1995. Managing financial reports of commercial banks: the influence of taxes, regulatory capital, and earnings. Journal of Accounting Research 33, 231–261.

Beaver, W.H., Kettler, P., Scholes, M., 1970. The association between market determined and accounting determined risk measures. Accounting Review 45, 654–682.

Beaver, W.H., McAnally, M.L., Stinson, C.H., 1997. The information content of earnings and prices: a simultaneous equations approach. Journal of Accounting and Economics 23, 53–81.

Bell, T.B., Landsman, W.R., Miller, B.L., Yeh, S., 2002. The valuation implications of employee stock option accounting for profitable computer software firms. The Accounting Review 77, 971–996.

Botosan, C., 1997. Disclosure level and the cost of equity capital. The Accounting Review 72, 323–349.

Botosan, C., Plumlee, M., 2002. A re-examination of disclosure level and expected cost of equity capital. Journal of Accounting Research 40, 21–40.

Bound, J., Jaeger, D.A., Baker, R.M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association 90, 443–450.

Brown, S., Hillegeist, S.A., Lo, K, 2004. Conference calls and information asymmetry. Journal of Accounting and Economics 37, 343–366.

Brown, S., Hillegeist, S.A., 2007. How disclosure quality affects the level information asymmetry. Review of Accounting Studies 12, 443–477.

Bushee, B.J., Matsumoto, D.A., Miller, G.S., 2003. Open versus closed conference calls: the determinants and effects of broadening access to disclosure. Journal of Accounting and Economics 34, 149–180.

Callen, J.L., Hope, O., Segal, D., 2005. Domestic and foreign earnings, stock return variability, and the impact of investor sophistication. Journal of Accounting Research 43, 377–412.

Chaney, P., Jeter, D., Shivakumar, L., 2008. Self-selection models and endogeneity issues in accounting research: the case of audit pricing. Working Paper, Vanderbilt University.

Claus, J.J., Thomas, J.K., 2001. Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets. Journal of Finance 56, 1629–1666.

Cohen, D.A., 2008. Does information risk really matter? An analysis of the determinants and economic consequences of financial reporting quality. Asia Pacific Journal of Accounting and Economics 15, 69–90.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. second ed Lawrence Earlbaum Associates, Hillsdale, NJ.

Copley, P.A., Gaver, J.J., Gaver, K.M., 1995. Simultaneous estimation of the supply and demand of differentiated audits: evidence from the municipal audit market. Journal of Accounting Research 33, 137–155.

Darrough, M., Rangan, S., 2005. Do insiders manipulate earnings when they sell their shares in an initial public offering? Journal of Accounting Research 43, 1–33.

DeFond, M.L., Raghunandan, K., Subramanyam, K.R., 2002. Do non-audit service fees impair auditor independence? Evidence from going concern audit opinions. Journal of Accounting Research 40, 1247–1274.

DiPrete, T.A., Gangl, M., 2004. Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. Sociological Methodology 34, 271–310.

D'Souza, J.M., 1998. Rate-regulated enterprises and mandated accounting changes: the case of electric utilities and post-retirement benefits other than pensions (SFAS No. 106). The Accounting Review 73, 387–410.

Francis, J., Lennox, C., 2008. Selection models in accounting research. Working Paper, University of Missouri at Columbia.

Frank, K.A., 2000. Impact of a confounding variable on a regression coefficient. Sociological Methods & Research 29, 147–194.

Frank, M.M., 2002. The impact of taxes on corporate defined benefit plan asset allocation. Journal of Accounting Research 40, 1163–1190.

Gebhardt, W.R., Lee, C.M.C., Swaminathan, B., 2001. Toward an implied cost of capital. Journal of Accounting Research 39, 135–177.

Gode, D.K., Mohanram, P., 2003. Inferring cost of capital using the Ohlson–Juettner model. Review of Accounting Studies 8, 399–431.

[15] The ability to make causal statements from a structural model does not, however, come without a cost. As noted, the specification of a structural econometric model requires the researcher to make a number of economic as well as statistical assumptions. The benefit of this exercise, however, is that the assumptions are made explicit and can be evaluated by other researchers.

Godfrey, L.G., Hutton, J.P., 1994. Discriminating between errors-in-variables/simultaneity and misspecification in linear regression models. Economic Letters 44, 359–364.
Greene, W.H., 2003. Econometric Analysis. fifth ed Prentice-Hall.
Guay, W.R., Kothari, S.P., Shu, S., 2003. Properties of implied cost of capital using analysts' forecasts. Working Paper, University of Pennsylvania.
Hahn, J., Hausman, J.A., 2003. Weak instruments: diagnoses and cures in empirical econometrics. American Economic Review 93, 118–125.
Hail, L., 2002. The impact of voluntary disclosure on the ex ante cost of capital for Swiss firms. European Accounting Review 11, 741–773.
Hall, A.R., Rudebusch, G.D., Wilcox, D.W., 1996. Judging instrument relevance in instrumental variables estimation. International Economic Review 37, 283–298.
Hamermesh, D.S., 2000. The craft of labormetrics. Industrial and Labor Relations Review 53, 363–380.
Hanlon, M., Rajgopal, S., Shevlin, T., 2003. Are executive stock options associated with future earnings? Journal of Accounting and Economics 36, 3–43.
Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271.
Haw, I., Hu, B., Hwang, L., Wu, W., 2004. Ultimate ownership, income management, and legal and extra-legal institutions. Journal of Accounting Research 42, 423–462.
Heckman, J.J., 1978. Dummy endogenous variables in a simultaneous equation system. Econometrica 46, 931–959.
Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica 47, 153–161.
Holthausen, R.W., Larcker, D.F., Sloan, R.G., 1995. Business unit innovation and the structure of executive compensation. Journal of Accounting and Economics 19, 279–313.
Hope, O., 2003. Disclosure practices, enforcement of accounting standards, and analysts' forecast accuracy: an international study. Journal of Accounting Research 41, 235–272.
Hunt, A., Moyer, S.E., Shevlin, T., 1996. Managing interacting accounting measures to meet multiple objectives: a study of LIFO firms. Journal of Accounting and Economics 21, 339–374.
Kang, S., Sivaramakrishnan, K., 1995. Issues in testing earnings management and an instrumental variable approach. Journal of Accounting Research 33, 353–367.
Kasznik, R., 1999. On the association between voluntary disclosure and earnings management. Journal of Accounting Research 37, 57–81.
Keating, A.S., 1997. Determinants of divisional performance evaluation practices. Journal of Accounting and Economics 24, 243–273.
Khurana, I.K., Raman, K.K., 2004. Litigation risk and the financial reporting credibility of Big 4 versus non-Big 4 audits: evidence from Anglo-American countries. The Accounting Review 79, 473–495.
Kothari, S.P., Zimmerman, J.L., 1995. Price and return models. Journal of Accounting and Economics 20, 155–192.
Lambert, R.A., Leuz, C., Verrecchia, R.E., 2007. Accounting information, disclosure, and cost of capital. Journal of Accounting Research 45, 385–420.
Lang, M.H., Lins, K.V., Miller, D.P., 2004. Concentrated control, analyst following, and valuation: do analysts matter most when investors are protected least? Journal of Accounting Research 42, 589–623.
Leuz, C., Verrecchia, R.E., 2000. The economic consequences of increased disclosure. Journal of Accounting Research 38, Supplement: Studies on Accounting Information and the Economics of the Firm, 91–124.
Lev, B., Sougiannis, T., 1996. The capitalization, amortization, and value-relevance of R&D. Journal of Accounting and Economics 21, 107–138.
Loudder, M.L., Khurana, I.K., Boatsman, J.R., 1996. Market valuation of regulatory assets in public utility firms. The Accounting Review 71, 357–373.
Lundholm, R.J., Myers, L.A., 2002. Bringing the future forward: the effect of disclosure on the returns–earnings relation. Journal of Accounting Research 40, 809–839.
Maddala, G.S., 1977. Econometrics. McGraw-Hill, New York.
Moreira, M.J., 2002. Tests with correct size in the simultaneous equations model. Ph.D. Thesis, University of California, Berkeley.
Moreira, M.J., Poi, B.P., 2001. Implementing tests with correct size in the simultaneous equation model. The Stata Journal 1, 1–15.
Murphy, K.J., 2000. Performance standards in incentive contracts. Journal of Accounting and Economics 30, 245–278.
Nagar, V., 2002. Delegation and incentive compensation. The Accounting Review 77, 379–395.
Nelson, C.R., Startz, R., 1990a. The distribution of the instrumental variables estimator and its *t*-ratio when the instrument is a poor one. Journal of Business 63, s125–s140.
Nelson, C.R., Startz, R., 1990b. Some further results on the exact small sample properties of the instrumental variable estimator. Econometrica 58, 967–976.
Ng, J., Rusticus, T.O., Verdi, R.S., 2008. Implications of transaction costs for the post earnings announcement drift. Journal of Accounting Research 46, 661–696.
Phillips, J.D., 2003. Corporate tax-planning effectiveness: the role of compensation-based incentives. The Accounting Review 78, 847–874.
Rajgopal, S., Shevlin, T., 2002. Empirical evidence on the relation between stock option compensation and risk taking. Journal of Accounting and Economics 33, 145–171.
Rajgopal, S., Venkatachalam, M., Kotha, S., 2003. The value relevance of network advantages: the case of e-commerce firms. Journal of Accounting Research 41, 135–162.
Reiss, P.C., Wolak, F.A., 2001. Structural econometric modeling: rationales and examples from industrial organization. Handbook of Econometrics 5.
Richardson, D.H., 1968. The exact distribution of a structural coefficient estimator. Journal of the American Statistical Association 63, 1214–1226.
Rosenbaum, P.R., 2002. Observational Studies. Springer, New York.
Rosenzweig, M.R., Wolpin, K.I., 2000. Natural "natural experiments" in economics. Journal of Economic Literature 38, 827–874.
Roulstone, D.T., 2003. The relation between insider-trading restrictions and executive compensation. Journal of Accounting Research 41, 525–551.
Sawa, T., 1969. The exact sampling distribution of ordinary least squares and two-stage least squares estimators. Journal of the American Statistical Association 64, 923–937.
Shi, C., 2003. On the trade-off between the future benefits and riskiness of R&D: a bondholders' perspective. Journal of Accounting and Economics 35, 227–254.
Shugan, S.M., 2004. Endogeneity in marketing decision models. Marketing Science 23, 1–3.
Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.
Stock, J.H., Wright, J.H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business & Economics Statistics 20, 518–529.
Weber, J., Willenborg, M., 2003. Do expert informational intermediaries add value? Evidence from auditors in microcap initial public offerings. Journal of Accounting Research 41, 681–720.
Whisenant, S., Sankaraguruswamy, S., Raghunandan, K., 2003. Evidence on the joint determination of audit and non-audit fees. Journal of Accounting Research 41, 721–744.
Willenborg, M., 1999. Empirical analysis of the economic demand for auditing in the initial public offerings market. Journal of Accounting Research 37, 225–238.
Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge.
Zivot, E., Startz, R., Nelson, C.R., 1998. Valid confidence intervals and inference in the presence of weak instruments. International Economic Review 39, 1119–1144.