Addressing unobserved endogeneity bias in accounting studies: control and sensitivity methods by variable type

Abstract

Together with their associated statistical routines, this paper describes the control and sensitivity methods that can be employed by accounting researchers to address the important issue of unobserved (omitted) variable bias in regression and matching models according to the types of variables employed. As with other social science disciplines, an important and pervasive issue in observational (non-experimental) accounting research is omitted variable bias (endogeneity). Causal inferences for endogenous explanatory variables are biased. This occurs in regression models where an unobserved (confounding) variable is correlated with both the dependent (outcome) variable in a regression model and the causal explanatory (often a selection) variable of interest. The Heckman treatment effect model has been widely employed to control for hidden bias for continuous outcomes and endogenous binary selection variables. However, in accounting studies, limited (categorical) dependent variables are a common feature and endogenous explanatory variables may be other than binary in nature. The purpose of this paper is to provide an overview of contemporary control methods, together with the statistical routines to implement them, which extend the Heckman approach to binary, multinomial, ordinal, count and percentile outcomes and to where endogenous variables take various forms. These contemporary methods aim to improve causal estimates by controlling for hidden bias, though at the price of increased complexity. A simpler approach is to conduct sensitivity analysis. This paper also presents a synopsis of a number of sensitivity techniques and their associated statistical routines which accounting researchers can employ routinely to appraise the vulnerability of causal effects to potential (simulated) unobserved bias when estimated with conventional regression and propensity score matching estimators.

Keywords: unobserved bias, control methods, sensitivity techniques, limited dependent variables, endogenous variable types, statistical routines

Addressing unobserved endogeneity bias in accounting studies: control and sensitivity methods by variable type

1. Introduction

Together with their associated statistical routines, this paper describes the control and sensitivity methods that can be employed by accounting researchers to address the important issue of omitted variable (hidden) bias in regression and matching models according to the types of dependent and explanatory variables employed. As with other disciplines, in observational (non-experimental) accounting research, causal inferences from regression models are biased when an unobserved variable is correlated with an explanatory (endogenous) variable and the outcome (dependent) variable. It is known as the omitted variable (edogeneity) problem and occurs when a variable that is excluded from the regression model is correlated with both the outcome variable Y and the causal explanatory (often a selection) variable X, such that inferences attributed to X are biased.

Peel and Makepeace (2012, p. 637) illustrate this in a study of big 4 auditor premiums by omitting a variable (corporate size) from their regression model for audit fees. Since corporate size is a principal determinant of (positively correlated with) both audit fees and the selection of big 4 auditors, when it is omitted from the regression model the big 4 premium increases by 207%. This is because the big 4 variable (X) is now (erroneously) also partially capturing the positive impact of (omitted) auditee size on audit fees (Y), thereby substantially inflating the big 4 premium[1]. This is a more severe example of the omitted variable problem, but does serve to highlight the problem. A further example relates to studies (e.g. Ittonen *et al*. 2013) which report that female auditors are associated with higher quality corporate financial reporting outcomes relative to their male counterparts. These quality differences (estimated via regression models), which are attributed to inherent female traits (such as diligence and risk tolerance), may be biased if control variables (e.g. age, experience and education) are omitted which are correlated with female auditors (relative to male ones) and with the outcome Y (the quality of financial reporting). Of course, the more completely specified a model is in terms of relevant explanatory (control) variables the less likely it is to be prone to unobserved bias.

However, it may be impractical/impossible to collect all potentially relevant control variables. For instance, archival audit fee studies do not control for the quality of internal audits and controls (Clatworthy *et al*. 2009). As hypothesised by Ireland and Lennox (2002), if such attributes are significantly associated with

[1] Note this principle underpins the use of multivariate regression models. Specifically, if we compare the mean audit fees of big 4 and non-big 4 auditees (univariate analysis), we find that big 4 clients incur substantially higher fees. This is factually correct, but it is uninformative regarding whether big 4 auditors charge an incremental premium - e.g. for conducting a higher quality audit - known as the treatment effect (below). Other relevant factors (such as client size and complexity) which determine both big 4 selection and fees must be controlled for in the regression model.

auditor selection and audit fees, then premium estimates will be biased. Importantly, an omitted variable may result in an underestimate of the causal effect for X. In the example for big 4 premiums above, if an unobserved variable is negatively associated with the selection of big 4 auditors and positively associated with audit fees, then the regression estimate of the big 4 premium will be biased downward[2]. Note also that omitting a variable from a regression model which is correlated with Y but which is uncorrelated with X will not affect the causal estimate for X. In summary, to bias the causal estimate for X, an unobserved variable must be significantly correlated with X and Y. The stronger the correlation, the greater is the bias.

As explained below, the techniques discussed in this paper which account for unobserved bias employ a first-step regression model where X is specified as the dependent variable. With control methods (below), the errors (unexplained variation in X) from this model are then used as a surrogate for omitted variables in the second-step outcome regression model for Y. Where Y is continuous and X is a binary choice variable, the Heckman treatment effect model (below) is widely used in accounting studies to control for unobserved selection bias (e.g. Dedman and Kausar 2012, Srinidhi *et al*. 2011, Wu *et al*. 2012). However, many accounting studies are concerned with an outcome variable which is categorical. For instance, Keasey and Short (1990) employ an ordered probit model to investigate the factors associated with perceived accounting burdens and Collis *et al*. (2004) use a logit model to examine the demand for company voluntary audits.

Categorical variables may be dichotomous (binary), unordered with more than two categories (multinomial) or ordered (e.g. ordinal ratings). They are normally estimated with binary, multinomial and ordered logit or probit regressions models respectively (Greene 2003) and are extensively employed in accounting research. In an early exposition of the methods used to estimate categorical models in accounting studies, Elliott and Kennedy (1988, p. 202) stress that many research issues involve limited dependent variables, including loan, bankruptcy, bond rating and takeover prediction, choice of accounting methods and accounting standards lobbying. In reviewing accounting studies with limited dependent variables appearing in 14 journals, Barniv and McDonald (1999, p .39) report that ‘the importance of these categorical techniques is demonstrated by the fact that at least 289 articles … used these techniques from 1989 through 1996’. Though noting that binary outcome models were more frequently used, Leclere (1999, p. 716) found that, of the 21 reviewed accounting studies which employed ordered or multinomial models, 76% had ordinal outcomes. An analysis of all papers appearing in journals with accounting in their titles between 2007 and 2012 on Google

[2] Other things equal, the premium would be underestimated by the equivalent of the overestimate.

Scholar revealed that 4,823 contained the word logit or probit, of which 18.1% (14.2%) included the additional term ordered (multinomial) respectively. Hence in accounting studies, logit/probit models are more frequently employed, though a substantial proportion use ordered or multinomial ones.

A prime aim of this paper is to provide a non-technical synopsis of the contemporary methods and associated statistical (mostly user-written) routines which can be employed by accounting researchers to control for unobserved bias for regression models with categorical (binary, multinomial and ordered) dependent variables, together with those that have count and quantile outcomes. As discussed below, models which control for endogeneity are more complex and exacting than their standard regression counterparts and require an additional instrumental variable for credible implementation. An alternative (or complementary) strategy is to employ sensitivity techniques. This paper also describes a range of sensitivity methods. The objective of these simpler techniques is to gauge the robustness of standard regression and propensity score matched causal estimates to potential (simulated) hidden bias and can be applied routinely in accounting research.

As at least partly evidenced by this paper's bibliography, the econometric and statistical literature relating to the omitted variable problem is vast and diffuse. Given this, it is perhaps unsurprising that there has been some confusion in the accounting literature (below) regarding the appropriate specification and application of the techniques, including the employment of valid instruments (see Larcker and Rusticus 2010, Tucker 2011, Lennox *et al*. 2012). The price of attempting to control for unobserved variable bias based on observable information is increased complexity, not least in the form of an additional instrumental variable. Therefore, the endogeneity correction methods described in this paper should not be viewed as silver bullets[3] for the omitted variable problem, rather caution is warranted in terms of their practical implementation (*ibid*). Nonetheless, endogeneity is a key and persistent empirical research issue, given standard regression parameters for an endogenous explanatory variable are biased and associated causal inferences may be erroneous (Tucker 2011). Knowledge of contemporary techniques for addressing such bias is therefore important, notwithstanding the associated increased complexity and practical implementation issues (below).

Though this paper aims to furnish accounting researchers with a concise and non-technical overview of the extant methods which address endogeneity bias, it supplies comprehensive source references, including those for bespoke statistical modules, nearly all of which are implemented via the user-friendly and popular Stata statistical package. Other than for two long-standing methods, implementation details are not included in

[3] The mechanical implementation of any statistical estimator without sufficient thought to theoretical considerations and to correctly specifying the proposed model is clearly ill advised, not least with regard to the methods discussed in this paper.

Stata or other manuals[4]. In consequence, illustrative statistical documentation, examples of the techniques applied in social science research studies[5] and supporting Stata Journal papers are described and referenced for the techniques and statistical modules[6] discussed. Given the extensive ground covered, the aim is to furnish accounting researchers with comprehensive source materials and examples, supported with a detailed bibliography, to facilitate the implementation and appropriate application of the methods.

Additional approaches to address omitted variable bias not covered in this paper are panel methods which are applicable where, as well as being available in cross-section, observations for subjects are accessible for two or more time periods (see Wooldridge 2010) and natural experiments, where an exogenous event (e.g. a change in policy or regulation) facilitates the estimation of causal effects in a similar fashion to randomised studies (see e.g. Lennox and Pittman 2011, Kinney and Shepardson 2011, for accounting examples). When estimating causal effects, research designs exploiting natural experiments offer a powerful methodology for circumventing unobserved bias, though the opportunity to implement such experiments is inherently limited.

For non-econometricians, the theoretical and methodological underpinnings of the various models described in this paper may appear daunting, but their foundations, together with the underlying concepts, are logical and relatively straightforward. Following an introduction to principal concepts, Section 2 describes the standard Heckman treatment effect model for hidden bias, together with recent modifications to accommodate non-binary explanatory variables. Key specification issues are also highlighted. The remainder of Section 2 extends the analysis to encompass methods which have been developed to address hidden bias where outcome variables are dichotomous, multinomial, ordered, count and percentile in form. Section 3 focuses on an array of sensitivity methods which aim to gauge the vulnerability of causal estimates to potential (simulated) hidden bias. Section 4 concludes the paper.

2. Methods for controlling for omitted variable bias

2.1 *Background*

Whether using archival and/or survey data in observational (non-experimental) accounting studies, endogeneity is a major issue since it results in biased causal estimates. An explanatory variable X is endogenous when it is

[4] *Treatreg* and *ivprobit* (Table 1) have built-in Stata commands and are supported by Stata manuals. As well as being user-friendly, a major feature of Stata is that experts in their fields produce dedicated user-written modules (Table 1).
[5] Where accounting studies are unavailable to illustrate the methods, applications in social science research are referenced and briefly described. These supplement the more technical statistical/econometric papers which are also referenced and described. It is hoped that they will be informative for researchers interested in implementing the techniques. Experience suggests that studying examples of the methods applied in extant empirical studies is fruitful.
[6] With a computer attached to the internet, the user-written Stata modules (commands) described in this paper can be easily accessed (including help documentation) and implemented in Stata when using a computer with internet access by simply typing *findit* followed by command names listed in tables 1 and 2.

significantly correlated with the error term (residuals) of the estimated regression model for Y (the dependent or outcome variable). Equivalently, this occurs where an unobserved variable is jointly correlated with X (the endogenous variable) and Y. It is commonly known as omitted variable bias and is also referred to as hidden or unobserved bias, endogenous treatment effects and selection bias. When Y is continuous and X is a binary selection variable, the Heckman treatment effect model is often applied in accounting research to test or control for unobserved bias. For instance, Dedman and Kausar (2012) examine the impact of voluntary audits (the treatment variable) on corporate credit scores (the outcome variable) and employ a Heckman treatment effect model (below) to account for hidden bias. They report (p. 415) that causal treatment estimates are robust to unobserved selection bias. Here the superior credit score attributed to voluntary audits is the treatment effect.

Endogeneity also arises where there is simultaneous causality bias. This occurs when X determines Y and Y determines X. For example, simultaneous causality bias has been reported and controlled for in a study examining the relationship between audit and consultancy fees (Whisenant *et al*. 2003). In this case, in a similar manner obtaining for omitted variable bias, standard regression estimates are biased. Where X and Y are both continuous, the standard approach to control for omitted variable or simultaneous causal bias is instrumental variables two-stage least squares regression (2SLS). To identify the second stage outcome equation, the 2SLS method requires a least one additional instrumental variable (IV) that is a significant determinant of X, but which is not directly and significantly correlated with Y. Specifically, as discussed below, the IV must be independent from Y other than via its correlation with X (e.g. Whisenant *et al*. 2003). With this method ordinary least squares (OLS) is employed to estimate both stages (models). In the first stage X is regressed on the IV together with the remaining explanatory variables (called covariates) used to determine Y. The predicted (fitted) values for X are then included in the second stage OLS regression (in place of X) together with the remaining covariates. This process effectively purges X of the endogeneity bias due to correlated errors[7].

It is important to distinguish between the IV, control function and maximum likelihood (correlated error adjustment) methods which are used to address omitted variable bias. Rather than the fitted values of X (as per IV 2SLS) being included in the outcome equation, Heckman and similar control function two-step approaches (below) employ the residuals (errors) from the first stage regression for X as an additional control variable in the outcome regression for Y, together with X and the other explanatory variables. In this context, note that the

[7] IV regression methods are extensively employed in economic research where simultaneous causality bias frequently features. The origins of the method can be traced as far back as 1928 in an exposition of the estimation of demand and supply elasticities (Stock and Trebbi, 2003).

actual values of X minus the fitted (predicted) values equal the residuals. Alternatively, and equivalent, fitted values plus the residuals equal actual values. In simple terms, IV estimators use fitted values of X (net of residuals) to control for bias in the outcome model, whereas control function methods employ X and the residuals for X to proxy for omitted confounding variables (see Wooldridge 2010, pp. 937-951 and Greene 2003, pp. 787-790, for a comparison of the methodologies). As discussed below, as with the Heckman two-step control function approach, maximum likelihood (ML) techniques adjust (control) for correlated errors, but contemporaneously estimate both steps as one system. To avoid confusion, hereafter control function and ML estimators are referred to in combination as control methods to differentiate them from IV (fitted value) ones.

Although less common than simultaneity bias, omitted variable bias may also obtain in accounting studies where X and Y are both continuous. For example, Jiao (2010, p. 2551) and Chen *et al*. (2012, p. 366) employ 2SLS to control for potential omitted variable bias when examining the relationship between shareholder welfare and corporate performance and unionisation rates and bond yields respectively. 2SLS can be implemented with Stata using the ivreg2 command. Comprehensive details regarding the implementation of *ivreg2* (including endogeneity tests) are provided by Baum *et al*. (2003, 2007).

Unlike IV estimators, technically, other than where mentioned below, control methods do not require an additional IV, since the outcome model is identified by the nonlinearity of the residuals from the first stage probit (or less commonly logit) model. However, as discussed below, to avoid or mitigate potential estimation problems (e.g. multicollinearity), where possible, an IV should be employed. Tests of whether there is significant evidence of endogeneity for IV estimators (Baum *et al*. 2003) and control methods (Cong and Drukker 2000) are included in the statistical routines described here. As discussed below, identifying a valid IV may prove problematic in terms of empirical application. This has at least partly motivated the development of the simpler sensitivity techniques discussed in Section 3 (DiPrete and Gangl 2004). Though based on the Heckman approach, rather than controlling for unobserved bias, these methods aim to asses the robustness of conventional regression and propensity score matched causal estimates to potential (simulated) confounders.

The next section describes the standard Heckman treatment model, applicable for continuous dependent variables where endogenous selection variables are binary, together with recent econometric generalisations to cases where potentially endogenous explanatory variables are multinomial or ordered in type. Specification issues are also highlighted. The analysis is then extended to consider control and IV techniques for outcomes which are dichotomous, multinomial, ordered, count and percentile in nature. In the order they are discussed,

Table 1 lists the methods by outcome and endogenous variable type, form of estimation and the associated statistical packages and commands (modules) to implement them.

Table 1 about here

2.2 *Continuous outcomes and specification issues*

This section describes the standard Heckman treatment effect model for binary selection variables together with its recent extension to multinomial and ordinal ones. Specification issues, which obtain to all the control methods listed in Table 1 and described below, are also addressed.

2.2.1 *Heckman treatment model*

The Heckman equations provide the foundation from which more recent control (correlated error adjustment) methods have been developed for various types of outcomes and endogenous explanatory variables. The standard Heckman two-step model for an endogenous binary selection variable D_i and a continuous dependent variable Y_i is:

Outcome equation $$Y_i = \delta D_i + \beta X_i + \varepsilon_{1i} \qquad i=1, \ldots ,N \tag{1}$$

Selection equation $$D_i^* = \theta Z_i + \varepsilon_{2i} \qquad i=1, \ldots ,N \tag{2}$$

D=1 when D*>0 and D=0 otherwise

This gives the following two-step model:

$$Y_i = \delta D_i + \beta(X_i) + \sigma_{12}\lambda_i + \upsilon_i \qquad i=1, \ldots ,N \tag{3}$$

$$\lambda_i = \lambda_{1i} = \frac{\phi(Z_i\theta)}{\Phi(Z_i\theta)} \text{ if } D_i\text{=1 and } \lambda_{0i} = -\frac{\phi(Z_i\theta)}{1-\Phi(Z_i\theta)} \text{ if } D_i\text{=0} \tag{4}$$

where X and Z are vectors of variables, β and θ are estimated parameters, ϕ and Φ are the normal density and cumulative distribution functions and λ (lambda) is the error correction term (the inverse Mills ratio, IMR), which is also referred to as the generalised probit residual. In (3) δ is the coefficient of the treatment effect after controlling for hidden bias ($\sigma_{12}\lambda_i$), with errors ε_{1i} and ε_{2i} normally distributed. In (2), θ denotes the estimated probit parameters for the vector of variables Z_i, which includes the covariates (X_i) and any further IVs.

The term $\sigma_{12}\lambda_i$ signifies that the estimated coefficient for λ_i is determined by the covariance (correlation) between ε_{1i} and ε_{2i}. For the control methods listed in Table 1 and discussed below, a statistically significant correlation (ρ) - which is denoted rho in statistical output - is indicative of endogeneity. The larger the magnitude of ρ the greater is the bias. Alternatively, if ρ is statistically insignificant then there is no evidence of endogeneity and the original standard model estimates are preferred. For two-step methods, the significance of the λ coefficient is equivalent to the significance of rho. For full information maximum likelihood estimators (below), the significance of rho is based on chi-square likelihood ratio tests. It follows that if ε_{1i} and ε_{2i} are positively (negatively) correlated (equivalently λ attracts a positive/negative coefficient), then the estimated treatment effect (D_i) will decrease (increase). Tests of whether the errors (residuals) are correlated are contained in all the statistical modules discussed below and listed in Table 1.

As already noted, in contrast to IV estimators (other than for the special case of a binary selection variable discussed below), since formal identification arises from distribution assumptions, an additional IV is not required to identify the selection effect in the outcome regression. This is because (3) is identified via the nonlinearity of the IMR (λ). However, the employment of an IV is highly desirable (below). First step selection effects are typically estimated with probit models because, like standard OLS ones, the errors are assumed to be normally distributed. Specifically, probit estimators employ the cumulative distribution function of the standard normal distribution as apposed to the logistic cumulative distribution function of the logit model. As stressed by Tucker (2010, p. 45), where a logit model is employed (below), the logistic error distribution must be transformed by the inverse standard normal function to comply with the normality assumption. The Heckman model can be estimated with the two-step method or simultaneously via full information maximum likelihood[8] (ML). As stated by Wooldridge (2010, p. 469), ML 'is generally the most efficient estimation procedure in the class of estimators that use information on the distribution of endogenous variables'. However, ML estimators tend to be more vulnerable to misspecification problems[9] (Greene 2003, p. 521).

Tucker (2010, p. 33) observes that the Heckman approach 'has been increasingly used in accounting and finance research in recent years'. An example of the application of Heckman's treatment effect model is provided in Leuz and Verrecchia's (2000) influential study[10] which investigates the impact of companies' reporting choice (IAS versus GAPP) on their cost of capital. More contemporary examples include Choi *et al*.

[8] Although more efficient, because ML methods jointly estimate the parameters (including ε_{1i} and ε_{2i}) they are more difficult to implement computationally than their two-step counterparts. Prior to the huge increase in computer power, two-step methods were sometimes preferred (particularly for large samples) on this basis (Cong and Drukker, 2000). The two estimators may produce similar results. For instance, Cong and Drukker (2000) report treatment variable coefficients of 1.26 (1.27) after controlling for selection bias with two-step (ML) estimators in an empirical example which illustrates the application of the Stata *treatreg* command.

[9] As noted by a Reviewer, the ML method sometimes suffers from non-convergence problems whereas the two-step method always results in convergence. This is more likely to be an issue where more complex multinomial specifications are employed, as shown in Table 1.

[10] The SSRN records 360 citations of Leuz and Verrecchia's (2000) paper.

(2008) who examine whether big 4 auditors charge fee premiums, Bi and Gregory (2011) who explore the relationship between choice of finance and abnormal merger returns, Srinidhi *et al*. (2011) who study the impact of the presence of a female board member on earnings quality, Bayar and Chemmanur (2012), who study the effect of exit choice (IPOs versus acquisitions) on firm valuation, Wu *et al*. (2012) who examine whether politically connected firms exhibit superior performance measures and Chou (2013) who investigates whether the receipt of a credit rating conveys information about firms' future earnings. Cong and Drukker (2000), explain in detail (with examples) how to implement both the two-step and ML versions of the Heckman treatment effect model with the Stata built-in *treatreg* command. For the most recent version (13) of Stata, *treatreg* has been updated to the *etregress* command.

Tucker (2010) provides a detailed and informative evaluation of Heckman selection models, concluding (p. 48) that ML estimators are preferred to their two-step counterparts on efficiency grounds. She also briefly describes (p. 45) how variations of the Heckman approach are feasible for different types of outcome and explanatory variables[11]. In particular, Tucker (2010) stresses that in some accounting studies the Heckman two-step model has been incorrectly applied to cases where selection and outcome variables are both binary. As described below, such an approach is known as 'forbidden regression', though efficient and consistent ML estimators have recently been formulated for all forms of explanatory and dependent variables (below).

2.2.2 *Extensions to multinomial and ordinal explanatory variables*

Although less common than binary ones, potentially endogenous choice variables in accounting research may be ordinal (e.g. based on questionnaire surveys) or represent three or more unordered (multinomial) categories. Recently, the Heckman treatment effect model for continuous dependent variables has been extended to these cases. Multinomial selection variables are specified as N-1 (the base case) binary variables. For instance, in Clatworthy and Peel's (2007) study, there are three binary variables representing big 4, mid-tier and smaller auditors. Their outcome model for audit fees contains dummy variables for big 4 and mid-tier auditors which are assessed relative to small auditors (the omitted base case). Other accounting examples of continuous Y with multinomial X include accounting choices (e.g. stock valuation) on the cost of debt (Ahmed *et al*. 2002) and the impact of different modes of entry into new markets on corporate performance (Pan *et al*. 1999).

Deb and Trivedi (2006a) extend the Heckman treatment effect method to the multinomial selection case. They specify models for treatment selection (a multinomial logit model) and Y which account for unobserved

[11] Tucker (2010, p. 44) also notes it is 'not advisable' to use probit or ordered probit outcome models with probit selection models to correct for bias.

selection bias via correlated model errors in an OLS outcome model as described above. To ensure convergence, parameters are estimated via maximum simulated likelihood[12] (Greene 2003, pp. 512-521). The method is implemented with the Stata *mtreatreg* command[13] (Deb 2009). Two informative examples of the application of *mtreatreg* are found in the studies of Vasquez (2011) who investigate the relationship between different types of water services selected and rental prices paid and Abreu *et al*. (2014) who examine the impact of UK graduates' migration strategies (four choices) on their subsequent earnings.

Accounting studies with continuous outcome variables may employ potentially endogenous explanatory variables which are ordered in nature, including ordinal selections and ratings (e.g. of internal control weaknesses, Joe *et al*. 2011). Other examples include the influence of audit tenure (on an ordinal scale) on audit fees (Copley et. al. 1994) and the impact of credit ratings on the cost of debt (Shaw 2012). In an award winning Stata Journal paper, Roodman (2011) has recently developed a comprehensive Stata statistical module which implements a number of ML estimators for a variety of models with endogenous regressors. Roodman (2011) stresses (p. 11) that, given the standard assumption of normally distributed errors (above), jointly estimated ML models (for endogenous X and outcome Y), which control for correlated errors provide efficient and consistent causal estimates for X (pp. 17-18). Amongst others (below), models with continuous outcomes and potentially endogenous ordinal ones are jointly estimated with ML employing OLS and ordered probit models. It is implemented with the *cmp* Stata command. Roodman (2011, 2013) provides a detailed exposition of the methodology underpinning *cmp* together with its implementation with Stata. Abrate *et al*. (2011) use *cmp* to contemporaneously estimate ordered probit and OLS models to account for endogeneity when studying the relationship between ordinal hotel ratings and prices.

2.2.3 *Specification issues*

The control methods for limited dependent variables discussed below follow similar specifications to those shown in equations (1) to (3), though with different combinations of regression models, depending on the nature of Y and X (binary, multinomial, ordinal, count or percentile). As stressed above, as with the Heckman treatment effect model, they are all estimated under the assumption of jointly normally distributed errors.

[12] Rather than ML, simulated ML is utilised where, amongst others, multinomial variables (either as outcomes or explanatory ones) are employed in models with endogenous variables. In such cases, estimation may involve integrals with high dimension and no closed form solutions, such that simulated ML is the only viable estimator (see Arias and Cox, 1999, for an informative discussion of the methodology).
[13] The methodology is also appropriate (below) where Y is dichotomous or count in type (Deb and Trivedi, 2006b).

Failure of this assumption may lead to biased and inconsistent parameter estimates[14]. Despite this, the extant literature is largely silent on the validity of the assumption and (to the author's knowledge) empirical studies simply assume that it holds. Although statistical packages (including Stata) include formal tests of whether the residuals of OLS regression models are normally distributed, testing whether the assumption holds for the errors of models with limited dependent variables has proved more exacting, largely because of the truncated nature of the outcome variables. Though specification tests for the normality assumption have been developed for models with limited dependent variables, including those for probit (Bera *et al*. 1984), multinomial probit (Murphy 2007, p. 399) and ordered probit (Glewwe 1997) ones, to the author's knowledge, no extant statistical package includes these tests - despite the widespread employment of categorical dependent variables in accounting and other disciplines. However, more recently, Wilde (2008) has provided a 'simple representation' of the Bera *et al* (1984) normality test for probit models together with statistical code[15] for its implementation.

As already noted, and except as specifically highlighted, though the control methods discussed below do not technically require an additional IV, extant accounting research suggests that credible implementation necessitates the employment of a valid IV to ensure robust identification[16] of the outcome model (Clatworthy et al. 2009, Larcker and Rusticus 2010, Lennox *et al*. 2012). Such a variable is theoretically appropriate in determining the endogenous explanatory variable X (and not merely spuriously correlated with X) in the first-step model, but is unrelated to Y in the second step outcome model other than by its association with X. With the arrows denoting the direction of causation, then the relationship is $IV \mapsto X \mapsto Y$. Note that if the IV is a significant determinant of Y in the outcome regression, this suggests it is not a valid instrument, with the implication being that it should instead be employed as an additional control variable (Lennox *et al*. 2012).

Notwithstanding the appropriate use of IVs in accounting research (e.g. Whisenant *et al*. 2003, Jurkus *et al*. 2011, Chou 2013, Ammann *et al*. 2013), the difficulty of obtaining valid ones is a limitation of endogeneity correction techniques. Of course, the simulation methods described in Section 3 do not require IVs and can be applied routinely in accounting studies. With some exceptions where the IV (fitted value) method is employed, in the remainder of this Section the Heckman control approach is extended to limited dependent variables.

[14] In a similar manner to the error terms in the Heckman treatment effect model, note that the standard Pearson correlation coefficient for assessing the degree of linear association between two variables also assumes that the variables are jointly (bivariate) normally distributed.
[15] Wilde's (2008) code is specified (p. 121) for implementation with the LIMDEP statistical package.
[16] Lennox *et al*. (2012) show that a number of studies do not use - or employ unsuitable - IVs, leading to a lack of robustness of reported empirical findings.

2.3 *Dichotomous (binary) outcomes*

As discussed above, dichotomous logit and probit models are frequently employed in accounting studies, with binary selection variables also being a common feature. Recent examples include Aldamen *et al*. (2012) who explore the relationship between audit committee characteristics and (binary) firm performance, Robinson *et al*. (2012) who examine whether board characteristics influence the likelihood of corporate failure, Clatworthy and Peel (2013) who study the impact of voluntary audits on financial statement errors and Dedman *et al*. (2014) who investigate the factors associated with the demand for voluntary audits. As stressed by Wooldridge (2010, p. 597), Imbens and Wooldridge (2007, p. 16) and Greene (2006, p. 1), for distributional reasons, it is inappropriate to use probit (binary, multinomial or ordinal) estimators for both selection and outcome models employing the IV fitted value estimator or the Heckman two-step control function approach (known as forbidden regression). In both cases inconsistent estimates would result. However, ML estimators can be employed to provide consistent and efficient parameters for endogenous variables not only for binary outcome models (Greene, 2009) but also for multinomial and ordered ones.

For binary selection and outcome variables, a longer standing solution advanced by Angrist (2001) is to apply IV 2SLS. Binary outcomes estimated with OLS are known as linear probability models (LPM). Although OLS estimators for LPMs are consistent (Wooldridge, 2010, p. 562), unlike their logit and probit counterparts, predictions may lie outside the 0 and 1 range of the dependent variable. However, as noted by Imbens and Wooldridge (2007, p. 16), the IV LPM method 'seems to provide good estimates of the average treatment effect in many applications'. Wooldridge (2010, p. 598) demonstrates the application of 2SLS with LPMs in a study of the labour force participation decision and specifies (pp. 939-941) an extension of the standard 2SLS LPM which provides more efficient treatment estimates (see also Cerulli, 2012, p. 10). Firstly, a probit model is estimated for the binary endogenous variable (X) as a function of the covariates and any additional instrument(s). The predicted probabilities (PP) from this model are then used as an instrument in the standard 2SLS procedure. Specifically, the OLS LPM is employed in the first stage to estimate fitted values of X as a function of the covariates and PP. Fitted values for X are then included in the 2SLS LPM outcome regression.

As noted by Wooldridge (2010, p. 940), as with Heckman control methods, technically, the identification of the outcome equation does not require an additional IV since PP is a non-linear (probit) function of the covariates. However, the employment of an IV is clearly desirable (above). Cerulli (2012) has written a Stata module (command: *ivtreatreg*) which implements standard LPM 2SLS and LPM 2SLS with a probit instrument as per Wooldridge (2010). Since linear probability models are prone (inherently) to heteroskedasticity, robust and bootstrap options (see Flachaire, 2005) are available with *ivtreatreg* to calculate appropriate standard errors. Cerulli (2012) provides a comprehensive exposition of these methods, including empirical examples.

The control methods discussed here for binary outcomes (as with those for other types of outcomes below), are analagous to the Heckman modelling approach described in Section 2.2. They employ the errors from the first-step models for X as surrogates for omitted variables in second-step models for Y as per Heckman and assume the errors are jointly normally distributed - the difference being in the type of estimators (e.g., for binary, multinomial and ordinal variables) employed in the first and second step equations (above), depending on the form X and Y take. As shown in Table 1, a range of Stata modules have been developed for binary outcomes where endogenous variables are binary, multinomial, ordinal or continuous in form. Other than when X is continuous (the reverse of the Heckman treatment effect model) when a two-step approach can be used, the methods employ ML to jointly estimate the first (endogenous variable) and second (outcome) models. Note also that if researchers are interested in comparing the coefficients (β) of models estimated with the logit, probit or LPM models described here, the following approximate relationships hold (Amemiya, 1981): $\beta_{probit} \approx 2.5\beta_{OLSLMP}$ and $\beta_{probit} \approx 0.625\beta_{logit}$.

As well as for ordinal and count outcomes (below), Miranda and Rabe-Hesketh (2006) have written a Stata module (command: *ssm*) which provides efficient and consistent estimates for binary outcomes with endogenous binary treatment variables via joint ML estimation of probit selection and outcome models. They provide comprehensive details of the methodology and examples of how *ssm* is implemented. They also demonstrate how, as an alternative to the probit outcome model, a logit specification may be employed. This is achieved (p. 288) by rescaling the logistic error distribution (above). Hassan *et al*. (2010) utilise *ssm* to control for selection bias when investigating the impact of health programme participation on the likelihood of being hospitalised and report similar causal inferences for logit and probit model specifications.

Using similar methodology to Miranda and Rabe-Hesketh (2006), Chiburis *et al*. (2011, 2012) have also developed a Stata module (command: *biprobittreat*) to implement ML estimation of dichotomous probit selection and outcome models. The module includes a bootstrap option which Chiburis *et al*. (2011) show, via Monte-Carlo simulations, may result in more robust standard errors, particularly in smaller samples. Brown *et al*. (2011) use *biprobittreat* with the bootstrap option when investigating whether the receipt of dental care is associated with the subsequent occurrence of cardiovascular disease.

The methodology[17] described in Section 2.2.2 for continuous Y and multinomial X is extended by Deb Trivedi (2006a) to where Y is binary and is implemented with the Stata *mtreatreg* module (Deb 2009). Simulated ML is employed to jointly estimate binary (multinomial) logit models for X (Y). Deb and Trivedi (2006a) provide a detailed explanation of the methodology together with empirical examples. In particular, normalisation parameters are employed (p. 311) to ensure consistent treatment effects are estimated. Zahabi *et al*. (2012) use *mtreatreg* when investigating the association between the choice of residential area (multinomial logit model) and travel mode (logit model). Similarly, Morescalchi (2011) estimates models with *mtreatreg* to control for hidden bias regarding the impact of a multi-valued selection variable (housing tenure) on the likelihood of being unemployed.

In accounting research, binary logit or probit models may include ordinal explanatory variables, particularly when data is collected via survey instruments. For example, Dedman and Lennox (2009) utilise ordinal explanatory variables reflecting perceived competition constructed from survey data in a probit model of the decision by medium-sized firms to file full or abbreviated accounts and Collis *et al*. (2004) employ a logit model which includes an ordinal variable measuring the perceived quality of information improvement as a potential determinant of voluntary audits. Given the standard assumption of normally distributed errors, Roodman's (2011) methodology (above) and his associated Stata *cmp* module is applicable to a range of models with differently measured X and Y variables. With regard to correlated model errors as described above, Roodman (2011, pp. 6-7) shows how efficient and consistent parameters are estimated with ML for endogenous ordinal variables with binary outcomes using ordered and binary probit models respectively. Marette *et al*. (2012), employ *cmp* to estimate the impact of an ordinal explanatory variable reflecting illness severity on the propensity to purchase a vaccine.

A long standing control function solution to the case where Y is binary and X is continuous is specified by Rivers and Vuong (1988). They formulate (pp. 352-353) a two-step procedure which produces consistent estimates where the residuals from a first step OLS model are included in a probit outcome model. As stressed by Rivers and Vuong (1988, p. 356), since identification does not rely on non-linearity, an additional IV must be included in the first step OLS model. Examples of accounting studies which employ Rivers and Vuong's (1988) methodology include Clatworthy and Peel (2013) who investigate how the proportion of women on boards affects the incidence of financial statement errors and Bagnoli *et al*. (2011) who examine whether

[17] The documentation describing how *mtreatreg* is implemented (Deb 2009) provides clear guidance of how Y is specified for logit, count and OLS models.

company interest rate spreads determine the likelihood of covenants appearing in debt contracts.

The Rivers and Vuong method is implemented with the Stata built-in *ivprobit* command either as a two-step procure or contemporaneously via ML. Wooldridge (2010, pp. 585-594) gives a detailed and instructive explanation of both estimators. Note that despite its name, rather than being an IV fitted value technique, *ivprobit* is a control method (above) adjusting for bias via correlated model errors (Wooldridge 2010, p. 587).

2.4 *Multinomial outcomes*

As detailed above, the focus of accounting studies is often a dependent variable with more than two unordered categories. For instance, De Cesari (2012) employs a multinomial logit model to examine the factors associated with 4 types of corporate dividend policy and Dimmock and Kouwenberg (2010) use a multinomial probit model to investigate the determinants of households' investment choices. Due to the difficulties in formulating an appropriate specification to account for correlated errors for selection and outcome equations, solutions for omitted variable bias for multinomial outcome models have, until recently, proved intractable. Though a natural extension of their binary counterparts, multinomial probit and logit estimators produce N-1 (base case) sets of model coefficients. For instance, a multinomial logit model is employed by Peel (1989) to examine the factors associated with firms which failed with and without going concern qualifications relative to non-failed ones. Model coefficients are evaluated with reference to non-failed companies (the omitted base case). As Imbens and Wooldridge (2007, p.17) note, the properties and specification of multinomial models have made it 'notoriously difficult' to specify an appropriate control function for unobserved bias.

Recently, however, Burgette and Nordheim (2009, 2010) have developed an estimator which employs a multinomial probit model for the outcome variable and a probit (multinomial probit) model for potentially endogenous selection variables which are binary (multinomial) in type respectively. The authors stress (2009, p. 2) that their method follows the Heckman treatment effect framework but utilises a Bayesian estimation methodology (Greene, 2006, pp. 429-427). Burgette and Nordheim (2009, 2010) provide comprehensive details of their estimator which is implemented with the R statistical package, command *endogMNP* (Burgette 2012). Both R and *endogMNP* are freely available[18]. Niankara (2011) uses *endogMNP* when investigating the association between health cost attitudes and health insurance choice.

Roodman's (2011) *cmp* Stata module (above) can also be utilised to control for endogenous regressors with multinomial dependent variables. Given normally distributed model errors (above), Roodman (2011, pp.

[18] The R package and R manuals can be downloaded from http://www.r-project.org/. The *endogMNP* module (and help files) can be downloaded from http://cran.r-project.org/web/packages/endogMNP/index.html.

7-10) explains how ML estimators provide efficient and consistent parameters employing multinomial probit outcome models and probit, multinomial probit, ordered probit and OLS ones for binary, multinomial, ordinal and continuous endogenous variables respectively. Employing *cmp,* De Paoli (2010, p. 27) examines the treatment effect associated with education attainment (binary) and subsequent fertility (multinomial) choices.

2.5 *Ordered outcomes*

As discussed above, after dichotomous outcomes, ordered ones appear to be the commonest limited dependent variable featuring in accounting studies. For instance, Allee and Yohn (2009) employ an ordered probit model when studying the determinants of financial statements use and Holmen and Pramborg (2006) examine the factors associated with the adoption of capital budgeting techniques via an ordered logit model. Like OLS regression - but unlike multinomial logit and probit models - one set of coefficients is estimated with ordered logit or probit models and interpreted in a similar fashion to OLS ones, though positive (negative) coefficients relate to the probability of being in higher (lower) categories of the ordinal dependent variable.

The *ssm* Stata module (above) can also be utilised to estimate ML models which account for endogeneity where outcome (explanatory) variables are ordinal (binary). Miranda and Rabe-Hesketh (2006, pp. 288-289) specify a similar formulation for correlated errors as described previously for binary outcomes. As before, a probit model is employed for the first step binary treatment variable with an ordered probit (or logit) one used for the outcome variable. Two informative applications of *ssm* to control for endogenous treatment effects with ordinal outcomes are the studies of Manasa (2009) who examine whether households who have an ‘elite’ member have greater access to public services (an ordered scale) and Flores-Fillol *et al*. (2010) who investigate the impact of firms’ payment (incentive) methods on employee cooperation, gauged on an ordinal scale.

Roodman’s (2011) *cmp* module (described above) can be used to control for endogeneity via ordered probit outcome models and multinomial probit, ordered probit and OLS first-step models for endogenous explanatory variables which are multinomial, ordinal and continuous in type respectively. Although to date there are no examples in the literature for the multinomial case, Vargas (2012) uses *cmp* when examining the relationship between an ordinal variable denoting firm size and an ordinal one reflecting perceptions of the obstacles firms face in achieving their objectives; whereas (Vargas 2012) employs *cmp* when studying whether individuals who have spent more time in education (a continuous variable) are more tolerant of homosexuals, as evaluated on an ordinal scale (Denny 2011).

2.6 *Count outcomes*

Count regression methods are formulated to account for the nature and distribution of cardinal non-negative dependent variables expressed as counts from zero upwards. Count outcomes (e.g. Greene 2006, pp. 740-747) are estimated via the Poisson distribution based on the number of occurrences of an event over a specified period. If the variance of the count variable is larger than its mean (known as overdispersion) then a modified poisson model is employed, known as negative binomial regression, which accounts for overdispersion. If there is no overdispersion, the negative binomial model reduces to the poisson model. Count models can be estimated with the Stata built-in *poisson* and *nbreg* commands[19].

Examples of count regression used in accounting research include Rock *et al*. (2001) who provide a detailed exposition of count regression methods with regard to the number of investment analysts following a firm and Dionne *et al*. (1996) who evaluate count estimators with regard to credit scoring systems (number of non-payments). More recent applications include Michels (2012) who investigates whether unverifiable disclosures influence the number of bids a loan listing receives, Cervellati *et al*. (2011) who study the relationship between personal investors' characteristics and number of stock trades and Weiss (2011) who examines whether firms' cost behavior influences the number of analysts following them.

To date, the Heckman treatment effect approach has been extended to count outcomes with binary and multinomial selection variables. Miranda (2004) and Miranda and Rabe-Hesketh (2006) describe how their ML methodology accounts for unobserved bias for binary selection variables via correlated errors (above) estimated with poisson outcome and probit selection models. It is implemented with their Stata user-written *ssm* module. As described by Miranda (2004, p. 42) and Miranda and Bratti (2006, p. 12), noteworthy is that their estimator accommodates overdispersion (above). Miranda and Bratti (2006) provide a detailed description of the application of *ssm* in a study investigating whether higher education participation leads to a reduction in daily cigarette consumption. Evans *et al*. (2011) also use *ssm* when examining the impact of environmental auditing on the incidence (count) of non-compliance with clean air regulations. The most recent (13) version of Stata includes a built-in (*etpoisson*) command which uses a similar methodology to that of *ssm* for count outcomes.

The binary treatment effect count model of Miranda and Rabe-Hesketh (2006) has been generalised by Deb and Trivedi (2006a, 2006b) to multinomial treatment variables. Employing simulated ML to jointly estimate multinomial logit selection and negative binomial outcome models, hidden bias is controlled for via

[19] The Stata *poisgof* command can be employed to test whether *nbreg* is preferable to *poisson*.

correlated errors of the two models. It is implemented with the user-written Stata *mtreatreg* module (Deb and Trivedi 2006b, Deb 2009). Deb and Trivedi (2006b, pp. 251-253) provide an informative exposition of the application of *mtreatreg* in a study of the relationship between the type of health insurance chosen by individuals and the number of doctor visits they make per annum. Buckley (2007) also employs *mtreatreg* when examining the association between the type of school attended by children and the number of hours their parents devote annually to school activities.

2.7 *Quantile regression*

Quantile regression (QR) extends OLS (conditional mean) analysis to provide model estimates for different conditional quantiles (percentiles) of the distribution (including the median) of a continuous dependent variable (Koenker and Hallock, 2001). QR is a robust estimator[20], in that it is less sensitive to skewness and outliers than its OLS counterpart (Greene 2006, p. 448, Jayaraman and Milbourn 2012, p. 23). Recent examples of the application of QR in accounting studies include Peel and Makepeace (2012) who estimate auditor premiums across quartiles (including the median) and inter-quartiles of the distribution of audit fees, Jayaraman and Milbourn (2012) who compare QR median and OLS regession estimates of the influence of stock liquidity on managerial compensation and Grace and Leverty (2010) who examine the impact of insurance regulation over a range of percentiles of companies' reserves.

Abadie *et al*. (2002) specify an IV (fitted value) two-stage treatment effects model for endogenous binary variables estimated with QR. A defining feature of the technique is that a binary instrumental variable is required to identify treatment parameters[21]. The method has been developed by Frolich and Melly (2010) for implementation with the Stata user-written *ivqte* module. First stage predicted probabilities are estimated with a logit model. These are then included in a second stage quantile regression model. Frolich and Melly (2010) furnish comprehensive details of the methodology together with examples of its implementation with *ivqte*. Cawley and Meyerhoefer (2012) use *ivqte* when investigating whether a high body mass is associated with higher medical costs over an array of its percentiles.

Finally, based on conditions for consistency and asymptotic normality, Lee (2007) formulates a two-step control function estimator (above) for continuous endogenous variables. Residuals from a first step OLS regression model (which must include an additional instrumental variable) are included in a QR outcome

[20] The quantile estimator is more robust to outliers in that OLS minimises the sum of the squares of the residuals, whereas quantile regression minimises the sum of absolute residuals, thus giving less weight to outliers (Wooldridge 2010, p. 450).
[21] Abadie *et al*. (2002) note (p. 426) that if only a non-binary instrumental variable is available it can be transformed into a binary one for identification purposes (see also note 26).

model. Andini (2010) provides a detailed and informative description of Lee's (2007) two-step method in a study which investigates the impact of the years spent in education on subsequent wages. Lee's (2007) estimator can be implemented with the user-written Stata *cqiv* module (Chernozhukov *et al*. 2011).

Table 2 about here

3. Sensitivity methods for hidden bias

3.1 *Background*

All statistical methods have limitations/assumptions. For instance, standard OLS estimators assume normally distributed errors and that there are no endogenous regressors. By definition the latter is more likely to hold if the model is well specified in terms of appropriate variables. Because the control methods in Section 2 use model errors (residuals) for the endogenous regressor as a surrogate for unobserved variables, the assumptions underpinning them are more exacting. Although Heckman was awarded the Noble Prize for his original insights, implementation of the estimators discussed in Section 2 are more demanding than standard ones. Fitted value (IV) techniques require an additional instrumental variable and it is best practice to employ one with control methods (Lennox *et al*. 2012). Though the methods in Section 2 aim to improve causal estimates by controlling for bias, the Heckman approach has been shown to lack robustness if no, or inappropriate, instruments are employed in accounting studies (above). However, in a recent simulation study, Guo and Fraser (2010, p. 296) demonstrate how, relative to OLS and matching estimators, the Heckman treatment effect model was 'robust to hidden bias. It was the only model that provided accurate estimation of the treatment effect'.

Sensitivity evaluation methods adopt a simpler modelling framework and do not require an additional IV. They aim to gauge how strong the impact of a potential confounding variable must be - via its combined impact on the potentially endogenous variable and the outcome one - to negate causal effects estimated with standard methods. Specifically, the vulnerability of estimated causal effects to potential hidden bias is appraised with reference to simulated omitted (confounding) variables. As stressed by Rosenbaum (1991, p. 901), 'the challenge is to say something useful and specific about the degree of evidence provided by the study, in particular the degree to which hidden biases are a plausible threat.' They may be employed as alternative or complementary methods to those described in Section 2. For instance, in examining the relationship between analysts' incentives to overweight corporate management guidance and corporate earnings forecasts, Feng and McVay (2010) use a Heckman treatment effect model (above) and Frank's (2000) sensitivity method (below).

The remainder of this Section provides a description of extant sensitivity techniques together with studies which have utilised them. Table 2 lists the methods by form of outcome, endogenous variable type and their associated statistical routines for multivariate regression models (Panel A) and propensity score matching estimators (Panel B). Other than one technique which is freely available as a formatted Excel spread-sheet, they are implemented with bespoke user-written Stata modules and are all predicated on the control function approach discussed in Section 2.

3.2 *Sensitivity methods for multivariate regression models*

Standard regression (e.g. logit or OLS) models assume that there is no unobserved variable bias impacting on causal estimates. This is known as the unconfoundedness, ignorability or conditional independence assumption (CIA). As discussed above, to bias parameters for an explanatory variable (X) an unobserved confounding variable (CX) must be significantly correlated with both X and the outcome variable Y. Based on partial product correlations, Frank (2000) develops a sensitivity method for assessing how large the product of the correlations (*r*) must be in linear regression models where Y is continuous to render the coefficient of X (based on its standard errors) statistically insignificant[22]; where X may be binary, ordinal, multinomial or continuous.

The product of two dependent correlations (PC1) is computed as: PC1 = $r_{XCX} \times r_{YCX}$, where r_{XCX} is the correlation coefficient between X and simulated CX and r_{YCX} is the correlation coefficient between Y and CX. After controlling for the remaining variables in the regression model, the computed value for PC1 is then the required degree of association (threshold) required for CX to render X statistically insignificant. Frank (2000, p. 172) illustrates how PC can be assessed with reference to other benchmarked control variables included in the regression model. Here the PC for a specified control variable (CV) is computed as if it is the confounding (omitted) variable associated with X. Specifically, PC2 = $r_{XCV} \times r_{YCV}$. PC1 can then be compared to PC2. For example, if PC1 (PC2) = 0.2 (0.1) we can say that the impact of treating CV as a confounding variable (PC2) is only 50% of that required by a hidden variable (PC1) to render X statistically insignificant (Frank 2000, footnote 13). Frank has developed a freely available formatted Excel spreadsheet[23] with instructions of how to enter data to compute the parameters previously described. Larcker and Rusticus (2010, p. 202) provide an informative accounting example of the application of Frank's sensitivity method with specific regard to benchmarked sensitivity parameters for the control variables included in their regression model, when

[22] Altonji *et al*. (2005) specify a similar sensitivity method when examining the impact of the type of school attended and subsequent education attainment. Employing probit selection models and probit and OLS outcome models they examine the sensitivity of treatment estimates to simulated unobserved bias with reference to the correlation (ρ) between the errors of the selection and outcome models as per the Heckman treatment effect equations as shown in Section 2.2.1 above.
[23] Frank's Excel formatted file (including instructions) is available from www.msu.edu/~kenfrank/research.htm.

investigating the impact of disclosure quality on bid-ask spreads. Feng and McVay (2010) also use Frank's technique when studying the relationship between a treatment variable based on company equity issues and the magnitude of analysts' forecast revisions. They report (p. 1636) that to nullify their findings would require an unobserved variable to have an impact 39% greater than any of their control variables.

With reference to empirical examples, Imbens (2003) formulates a sensitivity method for regression models where Y is continuous and X is a potentially endogenous binary selection variable modelled under the assumption of the logistic (logit model) distribution. As with the Rosenbaum bounds technique for matched treatment estimates (below), the method (pp.126-128) is analogous to the Heckman treatment effect model (above), in that the impact of a potential confounding variable (CX) is assessed with reference a logit selection model for X and to an OLS regression model for Y. Sensitivity analysis is based on a graphical representation of all values CX must have via its contemporaneous correlations with X and Y (where they intersect on the graph) to result in the statistical insignificance of X. As with Frank's (2000) technique, findings for CX can be benchmarked against other covariates in the regression model. Clarke (2009) provides a comprehensive exposition of the methodology including its application to covariate benchmarking. For instance, after controlling for other covariates, he demonstrates (p. 61) how one intersect point on the graphed relationship implies that a CX would have to explain concomitantly 20% of the variation in a logit selection model for the treatment variable (X) and 15% of the variation of Y for X to be rendered statistically insignificant.

Harada (2011) has substantially extended the original research of Imbens (2003) and developed a user-written 'generalized sensitivity analysis' (*gsa*) Stata module. As well as Imbens's (2003) logit specification, *gsa* implements sensitivity analysis for confounding variables for multivariate regression models where Y is continuous and for binary outcome models estimated with probit or logit models. Potential endogenous explanatory variables may be binary, multinomial[24], ordinal or continuous. As previously noted, the methodology (pp. 3-7) is similar to that of the Heckman treatment effects two-step procedure (above), in that simulated degrees of correlation between a confounding variable (CX) and X and between CX and Y are computed such that X becomes statistically insignificant. As described by Harada (2011, p. 8), different t-values can be stipulated (e.g. from the 0.01 to the 0.1 significance levels) when gauging the degree of correlation required for CX to nullify the causal estimate for X in the regression model for Y.

As noted above, analysis can also be conducted with reference to the impact the control variables in the

[24] As with Frank's (2000) method, for multinomial treatment variables, sensitivity analysis is conducted on each of the N-1 binary treatment variables included in the outcome model.

regression model have when they are treated as a CX. Helpful guidance for implementing *gsa* has been produced by Harada (2012a, 2012b). For instance, a simple command option enables the researcher to specify various model combinations: logit, probit or OLS outcome models with logit, probit or OLS ones for potentially endogenous regressors. Haranda (2011) gives a comprehensive illustration of *gsa* with regard to two empirical studies. Firstly, he investigates (pp. 8-14) the sensitivity of treatment estimates for a binary selection variable (job training) with a continuous outcome (wages); and secondly, he extends the analysis (pp. 14-20) - including with regard to covariates - to examine the relationship between a continuous explanatory variable (proportion ethnic population) and a continuous outcome representing political participation.

Further examples of the application of *gsa* include Bowen and Mo (2012) who study the sensitivity to hidden bias of regression estimates for a continuous explanatory variable (governor salaries) with a continuous outcome (tax burdens) and Grewal *et al*. (2012) who report sensitivity parameters for the estimated treatment effect associated with a binary variable (students with friends of high socioeconomic status) on a binary outcome (students' propensity to drop out). Results for a simulated confounding variable as well as benchmarked covariates lead the authors to conclude (p. 23) that 'our results are not sensitive to the unconfoundedness assumption, and hence, quite robust to endogeneity concerns about selection effects.'

3.2 *Propensity score matching and sensitivity methods*

Matching is an intuitive and logical method of controlling for observed bias, after which the significance of mean treatment effects are usually evaluated with univariate statistical tests. Only observations with similar observed characteristics (covariate values) are compared when estimating average treatment (selection) effects. Because exact covariate matching leads to the 'curse of dimensionality' (where matching closely on more than one attribute is usually impractical), the method of propensity score matching (PSM) is frequently (and increasingly) employed in accounting research (Tucker 2011, Peel and Makepeace 2012).

The seminal research of Rosenbaum and Rubin (1983) demonstrates that matching on one variable, propensity scores (selection probabilities) is equivalent to matching on each of the individual covariates. The method typically proceeds as follows. A logit or probit selection model which contains the control (matching) variables is estimated for the treatment variable (e.g. the big 4 binary variable in audit fee studies). The predicted values are then used for matching purposes. The commonest method is pairwise nearest neighbour (NN) matching, where treated observations are matched to (counterfactual) untreated ones (with or without replacement) with the closest predicted probabilities. To ensure close NN matching, a caliper may be employed which specifies the maximum difference in probabilities which constitutes an acceptable match. Relative to regression methods, the perceived advantages of PSM are that functional form or specification assumptions are not required and linear extrapolation beyond the common support (treated and untreated cases with similar attributes) is avoided.

Rosenbaum (2005) also demonstrates that matching may mitigate the impact of any hidden bias, concluding (p. 6) that ‘reducing heterogeneity reduces both sampling variability and sensitivity to unobserved bias - with less heterogeneity, larger biases would need to be present to explain away the same effect’. This is consistent with the simulation study of Guo and Fraser (2010) who report (p. 295) that PSM produced a more accurate estimate of the treatment effect than OLS regression in the presence of an unobserved confounding variable. The flip side of PSM is that information on non-matched observations is lost. Specifically, regression methods estimate the average treatment effect over the whole sample, whereas PSM estimates the average treatment effect of the treated (ATT).

Furthermore, after PSM, significant differences may remain (covariate imbalance) between the variable values of the matched treated and untreated samples. In this case, two approaches may be adopted. Firstly, standard multivariate regression methods are employed to estimate models in the matched sample (Ho *et al*. 2007); and secondly, double PSM (Rubin 2001) is applied - that is, the PSM process (above) is repeated again in its entirety on the first PSM matched sample. Lawrence *et al*. (2011) and Minutti-Meza (2013) estimate regression models in PSM samples when investigating whether big 4 auditor quality differentials are explained by client characteristics and auditor industry specialisation respectively. Clatworthy and Peel (2013) employ double PSM to eliminate covariate differences (remaining bias) when evaluating the impact of voluntary audits on accounting errors. Recently, Makepeace and Peel (2013) have developed a model to estimate PSM treatment effects adjusting for unobserved bias via the inclusion of inverse Mills ratios (IMRs) for treated and untreated observations in regression models estimated in the PSM samples. The IMRs are estimated in the full (unmatched) sample employing the Heckman (probit model) treatment model approach[25] (above). As well as E-Views, R, SAS and S-Plus, PSM can be implemented with the Stata *psmatch2* command which has a number of matching options, including kernel methods (Guo and Fraser 2010).

As with standard regression methods, PSM treatment effects are estimated under the assumption of

[25] The authors note (p. 2424) that their proposed model is preliminary in that further research is required regarding its statistical properties. It can be implemented in Stata employing *treatreg* (saving IMRs) and *psmatch2*.

unconfoundedness (the CIA). More recently, Rosenbaum (2005, 2010) has developed a sensitivity method - Rosenbaum bounds (RB) - which quantifies the potential impact of an unobserved variable on treatment effects estimated with PSM for both binary and continuous outcomes. The RB method (see DiPrete and Gangl 2004, Peel and Makepeace 2012) assumes that the selection odds of treatment are initially the same (no bias) for all matched subjects (j and k) with observed (matched covariate) characteristics (X). Given this, then:

$$\mathrm{Odds}(X) = \frac{\pi_x}{1-\pi_x} \tag{4}$$

Via a logistic distribution, potential departures from the CIA (no bias) are in the form of an odds ratio:

$$\frac{1}{\Gamma} \leq \frac{\mathrm{Odds}(X_j)}{\mathrm{Odds}(X_k)} =\leq \Gamma \tag{5}$$

The parameter Γ represents the differential selection odds of matched cases into treatment.

The exponential of a (log odds) logit model coefficient gives the odds ratio (Γ). Where Γ=1, the PSM treatment effect is assumed to be bias free and hence the logit selection coefficient = Ln(1) = 0. Higher values of Γ show the increasing impact a potential confounding variable (CX) exerts via its dual association with selection and outcome on the treatment effect (ATT). For instance, if Γ = 2 a CX doubles the odds of selection into treatment. Rosenbaum (2005, 2010) derives bounds on statistical confidence intervals for matched ATT estimates as Γ varies, thus defining a critical value of Γ at which the treatment effect is statistically insignificant. For example, if an RB critical Γ-value = 1.5 (odds of 1:1.5), this implies that a CX must increase the odds of selection into treatment (e.g. of a big 4 auditor in a premium study) by 50% and jointly exert a pro rata impact on Y such that the PSM estimated treatment estimate is statistically insignificant.

Note that RB parameters are conservative (worst case) in that they assume CX has an almost perfect association (for a given Γ-value) with Y. As stated by DiPrete and Gangl (2004, p. 278) it 'would almost perfectly predict which of a pair of matched cases would have the higher response'. For binary (continuous) outcomes RB bounds parameters are calculated with Mantel-Haenszel (Wilcoxon sign) statistics. Critical Γ-values can be specified for different (e.g. 0.01 to 0.1) significance levels (Clatworthy and Peel, 2013), and RB parameters may be generated for a CX which increases (as well decreases) the treatment effect (Peel and Makepeace, 2012). As explained and illustrated by DiPrete and Gangl (2004) and Peel and Makepeace (2012), CX critical Γ-values can be benchmarked against matching covariates ('hidden bias equivalents').

Two user-written Stata modules are available to implement RB. The first, *mhbounds* (Becker and

Caliendo 2007) is for binary outcomes, whereas the second, *rbounds* is for continuous ones (Gangl 2004). Both can be implemented after running the Stata *psmatch2* command (above). Peel and Makepeace (2012) employ *rbounds* to examine the vulnerability of PSM ATT estimates of auditor fee premiums to hidden bias, with Clatworthy and Peel (2013) using *mhbounds* when investigating the sensitivity of PSM treatment estimates for voluntary audits on the incidence of financial statement errors. Though current PSM and RB statistical packages do not facilitate simultaneous multinomial treatment comparisons, Peel and Makepeace (2012, p. 637) describe how this can be achieved using existing Stata modules to produce RB parameters for PSM premium differentials for three simultaneously matched auditor (big 4, mid-tier and small) categories.

Finally, based on the concurrent level of association between a confounding variable (CX) and the treatment and outcome variables, Ichino *et al*. (2008) have developed a sensitivity method to evaluate the vulnerability of PSM treatment estimates to hidden bias for binary selection[26] and outcome variables. The simulated CX is included as a matching variable with the other covariates. The PSM treatment effect (ATT) is then re-estimated to determine the impact of the CX. As explained by Ichino *et al*. (2008), as well as simulating a CX which renders the treatment effect statistically insignificant, the distribution of the simulated CX can be specified (benchmarked) to mirror that of observed matching covariates. A difference from RB is that it is not assumed that CX exactly predicts Y (for a given Γ-value).

Nannicini (2007) has written Stata module, *sensatt* to implement Ichino *et al*.'s technique and provides a comprehensive description and illustration of the method. Other informative examples of the application of *sensatt* include the studies of Millemaci and Sciulli (2011) who investigate the sensitivity (including covariate benchmarking) of PSM treatment estimates for a binary variable (representing childhood problems) and a binary outcome (employment status) and Loriga and Naticchioni (2010) who examine the sensitivity of the PSM ATT for job training schemes on the likelihood of obtaining employment. As well as results for a CX which nullifies the ATT, the authors report parameters for CXs which simulate the distributions of covariates used in the study.

4. Conclusion

In accounting and other social science research, potential hidden bias is a pervasive and perennial issue. For

[26] As explained by Nannicini (2007, p. 6), either in advance, or using *sensatt*, non-binary Y variables can be transformed to binary ones. For example, above and below the mean or median. This may be helpful to accounting researchers when Y is ordinal and can be readily partitioned into (say) high versus lower ratings.

instance, Clarke (2009) refers to unobserved bias as the ‘phantom menace’. This paper has provided an overview of extant control and sensitivity methods which accounting researchers can employ to address the important problem of hidden bias when estimating causal effects with regression and matching models. The classical Heckman treatment effect model is widely used in accounting studies to control for endogenous binary selection variables with continuous outcomes. However, the focus of accounting research is often a limited dependent variable and/or potentially endogenous variables that are non-binary in nature. As well as describing generalisations of the Heckman treatment effect model to multinomial and ordinal variables, this paper provides an overview of contemporary methods which extend the Heckman approach to models with binary multinomial, ordinal, count and quantile outcomes and to various types of endogenous variables.

Despite the econometric and statistical research effort expended to produce these methods, they are not a panacea for the hidden bias problem. As stressed in this paper and the extant accounting statistical literature (see Larcker and Rusticus 2010, Tucker 2011, Lennox *et al*. 2012), attempting to model (proxy for) an unobserved variable with respect to model errors for an endogenous explanatory variable comes at the cost of increased complexity, not least the requirement of an additional instrumental variable. Just as complexity increases from univariate analysis to standard multivariate methods, this is the price of attempting to model the impact of unobserved confounding variables from observable information. Specifically, it is well documented in accounting and other disciplines that standard control methods may lack robustness if no or inappropriate instruments are employed. Of course this does not imply that the control methods described in Section 2 lack utility. Rather that careful implementation is warranted[27] (Lennox *et al*. 2012). In particular, following Leamer (1983), Lennox *et al*. (2012, p. 610) advocate that researchers should evaluate the sensitivity (including to specification) of Heckman results relative to standard ones. A simpler and less statistically exacting approach is to use the sensitivity techniques described in Section 3 to assess the robustness of causal inferences to simulated (potential) omitted variable bias based on methodology analogous to that of Heckman treatment effect model. However, as stressed by Larcker and Rusticus (2010, p. 198) ‘There is no fool-proof way of of dealing of dealing with the problem of endogeneity in empirical accounting research’[28].

[27] Among a number of interesting observations, Lennox *et al*. (2012, p. 589) note ‘the frequent comments by editors and reviewers of the need to control for endogeneity’ - and that (p. 610) ‘Although OLS is typically more robust, it can still yield incorrect inferences when selection bias is a significant concern. Nevertheless, robustness is an important criterion that researchers should take into account when evaluating their findings’.

[28] If the researcher knows (e.g. with reference to prior research) or suspects a variable of potential import is omitted (e.g. because it is unavailable in an archival database), and has expectations regarding its likely impact, then the plausibility of control method causal estimates or sensitivity technique evaluations may be easier to assess.

Other things equal[29], accounting researchers may adopt three strategies: (i) assume the model is well specified in terms of explanatory variables (the CIA holds) and employ conventional estimators; (ii) apply the estimators described in Section 2 to control (or test) for omitted bias; or (iii) employ the sensitivity techniques described in Section 3 to appraise the vulnerability of standard causal estimates to confoundedness. It is hoped that this paper will be instrumental in facilitating implementation by accounting researchers of (ii) and (iii) for a range of model specifications with different types of outcome and explanatory variables. Of course these options are not mutually exclusive. For instance, if applying the control methods in Section 2 proves impractical, then the sensitivity ones in Section 3 can be applied. Given that sensitivity techniques can be implemented free of the complexities associated with edogeneity correction methods, their routine application to appraise the robustness of standard regression and matching estimators to hidden bias in accounting studies is clearly desirable. As concluded by DiPrete and Gangl (2004, p. 303), sensitivity analysis is 'an important tool for assessing the level of caution that one should use when interpreting the significance tests for causal effects that are produced with conventional estimators'.

[29] As highlighted in Section 1, the exploitation of natural experiments offers a powerful methodology for addressing endogeneity concerns in accounting studies (Gassen 2013). As also noted in the Introduction, additional methods for dealing with endogeneity are available where studies employ panel data (Wooldridge 2010).

References

Abadie, A., Angrist, J. and Imbens, G., 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70 (1), 91–117.

Abrate, G., Capriello, A. and Fraquelli, G., 2011. When quality signals talk: evidence from the Turin hotel industry. *Tourism Management*, 32 (4), 912–921.

Abreu, M., Faggian, A. and McCann, P., 2014. Migration and inter-industry mobility of UK graduates. *Journal of Economic Geography*. Early (online), 29 January 2014. Available from: http://joeg.oxfordjournals.org/content/early/2014/01/29/jeg.lbt043.abstract [Accessed 3 May 2014].

Ahmed, A., Billings, B., Morton, R. and Harris, M., 2002. The role of accounting conservatism in mitigating bondholder-shareholder conflicts over dividend policy and in reducing debt costs. *Accounting Review*, 77 (4), 867–890.

Aldamen, H., Duncan, K., Kelly, S., McNamara, R. and Nagel, S., 2012. Audit committee characteristics and firm performance during the global financial crisis. *Accounting and Finance*, 52 (4), 971–1000.

Allee, K. and Yohn, T., 2009. The demand for financial statements in an unregulated environment: an examination of the production and use of financial statements by privately-held small businesses'. *Accounting Review*, 84 (1), 1–25.

Altonji, J., Todd, E., Elder, T. and Taber, C., 2005. Selection on observed and unobserved variables: assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113(1), 151–184.

Amemiya, T., 1981. Qualitative response models: a survey. *Journal of Economic Literature*, 19 (4), 1483–1536.

Ammann, M., Hoechle, D. and Schmid, M., 2013. Is there really no conglomerate discount? *Journal of Business Finance & Accounting*, 39 (1&2), 264–288.

Andini, C., 2010. Within-groups wage inequality and schooling: further evidence for Portugal. *Applied Economics*, 42 (28), 3685–3691.

Angrist, J., 2001. Estimation of limited endogenous variable models with dummy endogenous regressors: simple strategies for empirical practice. *Journal of Business and Economic Statistics*, 19 (1), 2–16.

Arias, C. and Cox, T., 1999. Maximum simulated likelihood: a brief introduction for practitioners.Working paper. Available from: www.child-centre.unito.it/papers/child07_2003.pdf [Accessed 1 March 2013].

Bagnoli, M., Liu, H. and Watts, S., 2011. Family firms, debtholder-shareholder agency costs and the use of covenants in private debt. *Annals of Finance*, 7 (4), 477–509.

Barniv, R. and McDonald, J., 1999. Review of categorical models for classification issues in accounting and finance. *Review of Quantitative Finance and Accounting*, 13 (1), 39–62.

Baum, C., Schaffer, M. and Stillman, S., 2003. Instrumental variables and GMM: estimation and testing. *Stata Journal*, 3 (1), 1–31.

Baum, C., Schaffer, M. and Stillman, S., 2007. Enhanced routines for instrumental generalized method of moments estimation and testing. *Stata Journal*, 7 (4), 465–506.

Bayar, O. and Chemmanur, T., 2012. What drives the valuation premium in IPOs versus acquisitions? An empirical analysis. *Journal of Corporate Finance,* 18 (3), 451–475.

Becker, S. and Caliendo, M., 2007. Sensitivity analysis for average treatment effects. *Stata Journal*, 7 (1), 71–83.

Bera, A., Jarque, C. and Lee, L., 1984. Testing the normality assumption in limited dependent variable models. *International Economic Review*, 25 (3), 563–578.

Bi, X. and Gregory, A., 2011. Stock market driven acquisitions versus the Q theory of takeovers: the UK evidence. *Journal of Business Finance & Accounting*, 38 (5&6), 628–656.

Bowen, T. and Mo, C., 2012. The voter's blunt tool. Working paper. Available from: http://ssrn.com/abstract=2135011[Accessed 4 April 2013].

Brown, T., Cruz, E. and Brown, S., 2011. The effect of dental care on cardiovascular disease outcomes: an application of instrumental variables in the presence of heterogeneity and self-selection. *Health Economics*, 20 (10), 1241–1256.

Buckley, J., 2007. Choosing schools, building communities? The effect of schools of choice on parental involvement. Working paper. Available from: http://ncspe.org/publications_files/OP133.pdf [Accessed 23 May 2013].

Burgette, L. and Nordheim, E., 2009. A full Gibbs sampler for a multinomial probit model with endogeneity. Working paper. Available from: http://www.stat.duke.edu/~lb131/switchBurgetteNordheim.pdf [Accessed 25 March 2013].

Burgette, L. and Nordheim, E., 2010. A full Gibbs sampler for the Bayesian multinomial probit switching model. Working paper. Available from: http://www.burgette.org/switchBurgetteNordheim.pdf [Accessed 25 March 2013].

Burgette, L. and Nordheim, E., 2010. A full Gibbs sampler for the Bayesian multinomial probit switching model. Working paper. Available from: http://www.burgette.org/switchBurgetteNordheim.pdf [Accessed 25 March 2013].

Burgette, L., 2012. Package: endogMNP. Available from: http://cran.r-project.org/web/packages/endogMNP/endogMNP.pdf [Accessed 25 March 2013].

Cawley, J. and Meyerhoefer, C., 2012. The medical care costs of obesity: an instrumental variables approach. *Journal of Health Economics*, 31 (1), 219–230.

Cerulli, G., 2012. Ivtreatreg: a new Stata routine for estimating binary treatment models with heterogeneous response to treatment under observable and unobservable selection.Working paper. Available from: http://www.ceris.cnr.it/ceris/workingpaper/2012/WP_3_CERULLI.pdf [Accessed 3 January 2013].

Cervellati, E., Fattori, P. and Pattitoni, P., 2011. Individual investor behaviour: evidence from the clients of a small credit cooperative bank. *International Journal of Behavioural Accounting and Finance*, 2 (3&4), 191–207.

Chen, H., Kacperczyk, M. and Ortiz-Molina, H., 2012. Do nonfinancial stakeholders affect the pricing of risky debt? Evidence from unionized workers. *Review of Finance*, 16 (2), 347–383.

Chernozhukov, V., Fernandez-Val, I. and Kowalski, A., 2011. CQIV: Stata module to perform censored quantile instrumental variables regression. Available from: http://ideas.repec.org/c/boc/bocode/s457478.html [Accessed 19 April 2013].

Chiburis, R., Das, J. and Lokshin, M., 2011. A practical comparison of the bivariate probit and linear IV estimators. Working paper. Available from: http://elibrary.worldbank.org/content/serial/1813-9450 [Accessed 6 January 2013].

Chiburis, R., Das, J. and Lokshin, M., 2012. A practical comparison of the bivariate probit and linear IV estimators. *Economics Letters*, 117(3), 762–766.

Choi, J., Kim, J., Liu, X. and Simunic, D., 2008. Audit pricing, legal liability regimes and big 4 premiums: theory and cross-country evidence. *Contemporary Accounting Research*, 25 (1), 55–99.

Chou, T., 2013. Information content of credit ratings in pricing of future earnings. *Review of Quantitative Finance and Accounting*, 40 (2), 217–250.

Clarke, K., 2009. Return of the phantom menace: omitted variable bias in political research. *Conflict Management and Peace Science*, 26 (1), 46–66.

Clatworthy, M., Makepeace, G. and Peel, M., 2009. Selection bias and the big four premium: new evidence using Heckman and matching models. *Accounting and Business Research*, 39 (2), 207–233.

Clatworthy, M. and Peel, M., 2007. The effect of corporate status on external audit fees: evidence from the UK. *Journal of Business Finance & Accounting*, 34 (1&2), 169–201.

Clatworthy, M. and Peel, M., 2013. The impact of voluntary audits and governance characteristics on accounting errors in private companies. *Journal of Accounting and Public Policy*, 32 (3), 1–25.

Collis, J., Jarvis, R. and Skerratt, L. 2004. The demand for the audit in small companies in the UK. *Accounting and Business Research*, 34 (2), 87–100.

Cong, R. and Drukker, D., 2000. Treatment effect model. *Stata Technical Bulletin*, 55 (May), 25–33.

Copley, P., Doucet, M. and Gaver, K., 1994. A simultaneous equations analysis of quality control review outcomes and engagement fees for audits of recipients of federal financial assistance. *Accounting Review*, 69 (1), 244–256.

De Cesari, A. 2012. Expropriation of minority shareholders and payout policy. *British Accounting Review*, 44 (4), 207–220.

De Paoli, A., 2010. The effect of schooling on fertility, labor market participation and childrens outcomes: evidence from Ecuador. Working paper. Available from: http://econpapers.repec.org/paper/milwpdepa/2010-30.htm [Accessed 8 March 2013].

Deb, P., 2009. Mtreatreg: Stata module to fit models with multinomial treatments and continuous, count and binary outcomes using maximum simulated likelihood. Available from: http://ideas.repec.org/c/boc/bocode/s457064.html [Accessed 7 January 2013].

Deb, P. and Trivedi, P., 2006a. Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: application to health care utilization. *Econometrics Journal*, 9 (2), 307–331.

Deb, P. and Trivedi, P., 2006b. Maximum simulated likelihood estimation of a negative-binomial regression model with multinomial endogenous treatment. *Stata Journal*, 6 (2), 246–255.

Dedman, E. and Kausar, A., 2012. The impact of voluntary audit on credit ratings: evidence from UK private firms. *Accounting and Business Research*, 42 (4), 397–418.

Dedman, E., Kausar, A. and Lennox, C., 2014. The demand for audit in private firms: recent large-sample evidence from the UK. *European Accounting Review*, 23 (1), 1–23.

Dedman, E. and Lennox C., 2009. Perceived competition, profitability and the withholding of information about sales and the cost of sales. *Journal of Accounting and Economics*, 48 (2&3), 210–230.

Denny, K., 2011. Civic returns to education: its effect on homophobia. Working paper. Available from: http://ideas.repec.org/p/ucd/wpaper/201109.html [Accessed 4 March 2013].

Dimmock, S. and Kouwenberg, R., 2010. Loss-aversion and household portfolio choice. *Journal of Empirical Finance*, 17 (3), 441–459.

Dionne, G., Artis, M. and Guillen, M., 1996. Count data model for a credit scoring system. *Journal of Empirical Finance*, 3 (3), 303–325.

DiPrete, T. and 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 (1), 271–310.

Elliott, J. and Kennedy, D., 1988. Estimation and prediction of categorical models in accounting research. *Journal of Accounting Literature*, 7, 202–242.

Evans, M., Liu, L. and Stafford, S., 2011. Can facilities police themselves? Evidence on the effectiveness of environmental auditing.Working paper. Available from: http://www.webmeets.com/files/papers/aere/2011/153/Can%20Facilities%20Police%20Themselves%20May%202011.pdf [Accessed 4 March 2013].

Feng, M. and McVay, S., 2010. Analysts incentives to overweight management guidance when revising their short-term earnings forecasts. *Accounting Review*, 85 (5), 1617–1646.

Flachaire, E. 2005. Bootstrapping heteroskedastic regression models: Wild bootstrap vs. pairs bootstrap. *Computational Statistics and Data Analysis*, 49 (2), 361–376.

Flores-Fillol, R., Iranzo, S. and Ferran, M., 2010. Workers cooperation within the firm: an analysis using small and medium size firms. Working paper. Available from: http://www.recercat.cat//handle/2072/179606 [Accessed 17 March 2013].

Frank, K., 2000. Impact of a confounding variable on a regression coefficient. *Sociological Methods and Research*, 29 (2), 147–194.

Frolich, M. and Melly, D., 2010. Estimation of quantile treatment effects with Stata. *The Stata Journal*, 10 (3), 423–457.

Gangl, M., 2004. Rbounds: Stata module to perform Rosenbaum sensitivity analysis for average treatment effects on the treated. Available from: http://ideas.repec.org/c/boc/bocode/s438301.html [Accessed 30 March 2013].

Gassen, J. (2013). Causal inference in empirical archival financial accounting research. *Accounting, Organizations and Society*, Early (online), 14 Mar 2013. Available from: http://www.sciencedirect.com/science/article/pii/S0361368213000706 [Accessed 3 May 2014].

Glewwe, P., 1997. A test of the normality assumption in the ordered probit model. *Econometric Reviews*,16 (1), 1–19.

Grace, M. and Leverty, J. 2010. Political cost incentives for managing the property-liability insurer loss reserve. *Journal of Accounting Research*, 48 (1), 21–49.

Greene, W., 2003. *Econometric Analysis*. 5th ed. New York: Prentice-Hall International.

Greene, W., 2006. A general approach to incorporating selectivity in a model. Working paper. Available from: http://ideas.repec.org/p/ste/nystbu/06-10.html [Accessed 25 January 2013].

Greene, W., 2009. Discrete choice modeling. In: T. Mills and K. Patterson, eds. *The Handbook Econometrics*. London: Palgrave, pp. 473–556.

Grewal, E., Mo, C., Williams, B. and Nie, N., 2012. With a little help from my friends' parents: exploring peer effects and educational attainment. Working paper. Available from: http://www.stanford.edu/~etgrewal/Grewaletal_SES_Attainment_2012 [Accessed 11 January 2013].

Guo, S. and Fraser, M., 2010. *Propensity Score Analysis: Statistical Methods and Applications*. London: Sage Publications.

Harada, M., 2011. Generalized sensitivity analysis.Working paper. Available from: https://files.nyu.edu/mh166/public/docs/research.html#gsa [Accessed 11 January 2013].

Harada, M., 2012a. GSA: Stata module to perform generalized sensitivity analysis. Available from: http://ideas.repec.org/c/boc/bocode/s457497.html [Accessed 2 April 2013].

Harada, M., 2012b. GSA: generalized sensitivity analysis. Available from: http://fmwww.bc.edu/repec/bocode/g/gsa.html [Accessed 2 April 2013].

Hassan, A., Jimenez, J. and Montoya, R., 2010. The impact of subsidized health insurance on the poor in Colombia: evaluating the case of medellin through endogenous switching models and propensity score matching for medical care utilization. Working paper. Available from: http://ssrn.com/abstract=2193826 [Accessed 9 March 2013].

Ho, D., Kosuke, I., King, G. and Stuart, E., 2007. Matching as nonparametric preprocessing for improving parametric causal inference. *Political Analysis*, 15 (3), 199–236.

Holmen, M. and Pramborg, B., 2006. Capital budgeting and political risk: empirical evidence. *Journal of International Financial Management and Accounting*, 20 (2), 105–134.

Ichino A., Mealli, F. and Nannicini, T., 2008. From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics*, 23 (3), 305–327.

Imbens, G., 2003. Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review*, 93 (2), 126–132.

Imbens, G. and Wooldridge, J., 2007. *What's New in Econometrics*? Massachusetts: National Bureau of Economic Research.

Ireland, J. and Lennox C., 2002. The large audit firm fee premium: a case of selectivity bias?, *Journal of Accounting, Auditing and Finance*, 17 (1), pp. 73–91.

Ittonen, K., Vahamaa, E. and Vahamaa, S., 2013. Female auditors and accruals quality. *Accounting Horizons*, 27 (2), 205–228.

Jayaraman, S. and Milbourn, T., 2012. The role of stock liquidity in executive compensation .*Accounting Review*, 87 (2), 537–563.

Jiao, Y., 2010. Stakeholder welfare and firm value. *Journal of Banking and Finance*. 34 (10), 2549–2561.

Joe, J., Wright A. and Wright, S., 2011. The impact of client and misstatement characteristics on the disposition off proposed audit adjustments. *Auditing: A Journal of Practice & Theory*, 30 (2), 103–124.

Jurkus, A., Park, J. and Woodard, L., 2011. Women in top management and agency costs. *Journal of Business Research*, 64 (2) 180–196.

Keasey, K. and Short, H., 1990. The accounting burdens facing small firms: an empirical research note. *Accounting and Business Research*, 20 (80), 307–313.

Kinney, W. and Shepardson, M., 2011. Do control effectiveness disclosures require SOX 404(b) internal control audits? A natural experiment with small U.S. public companies. *Journal of Accounting Research*, 49 (2), 413–448.

Koenker, R. and Hallock, K., 2001. Quantile regression, *Journal of Economic Perspectives*, 15 (4), 143–156.

Larcker, D. and Rusticus, T., 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics*, 49 (3), 186–205.

Lawrence, A., Minutti-Meza, M. and Zhang, P., 2011. Can big 4 versus non-big 4 differences in audit quality proxies be attributed to client characteristics? *Accounting Review*, 86 (1), 259–286.

Leamer, E., 1983. Lets take the con out of econometrics. *American Economic Review*, 73, (1), pp. 31–43.

Leclere, M., 1999. The interpretation of coefficients in models with n-qualitative or ordered dependent variables. *Contemporary Accounting Research*, 16 (4), 711–747.

Lee, S., 2007. Endogeneity in quantile regression models: a control function approach. *Journal of Econometrics*, 141 (2), 1131-1158.

Lennox, C., Francis, J. and Wang, Z., 2012. Selection models in accounting research. *Accounting Review*, 87 (2), 589–616.

Lennox, C. and Pittman, J., 2011. Voluntary audits versus mandatory audits. *Accounting Review*, 86 (5), 1655–1678.

Leuz, C. and Verrecchia, R., 2000. The economic consequences of increased disclosure. *Journal of Accounting Research*, 38 (supplement), 91–124.

Loriga, S. and Naticchioni, P., 2010. Short and long term evaluations of public employment services in Italy. Working paper. Available from: http://ssrn.com/abstract=1548641 [Accessed 4 April 2013].

Makepeace, G. and Peel, M., 2013. Combining information from Heckman and matching estimators: testing and controlling for hidden bias. *Economics Bulletin*, 33(3), 2422–2436.

Manasa, P., 2009. Power, politics and public good provision: understanding institutional persistence in rural Punjab. Working paper. Available from: https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=res2010&paper_id=216 [Accessed 15 March 2013].

Marette, S., Roe, B. and Teisl, M., 2012. The welfare impact of food pathogen vaccines. *Food Policy*, 37 (1), 86–93.

Michels, J., 2012. Do unverifiable disclosures matter? evidence from peer-to-peer lending. *Accounting Review*, 87 (4), 1385–1413.

Millemaci, E. and Sciulli, D., 2011. The causal effect of family difficulties during childhood on adult labour market outcomes. Working paper. Available from: http://ssrn.com/abstract=1883433 [Accessed 2 April 2013].

Minutti-Meza, M., 2013. Does auditor industry specialization improve audit quality? *Journal of Accounting Research*, 51(4), 779–817.

Miranda A., 2004. FIML estimation of an endogenous switching model for count data. *Stata Journal*, 4 (1), 40–49.

Miranda, A. and Bratti, M., 2006. Non-pecuniary returns to higher education: the effect on smoking intensity in the UK, Working paper. Available from: http://hdl.handle.net/10419/33228 [Accessed 20 March 2013].

Miranda, A. and Rabe-Hesketh, S., 2006. Maximum likelihood estimation of endogenous switching and sample selection models for binary, count, and ordinal variables. *Stata Journal*, 6 (3), 285–308.

Morescalchi, A., 2011. Housing tenure and individual labour market outcomes: an empirical assessment based on the UK Labour Force Survey. Working paper. Available from: www.aiel.it/bacheca/milano/papers/Morescalchi2.pdf [Accessed 6 May 2013].

Murphy, A., 2007. Score tests of normality in bivariate probit models. *Economics Letters*, 95 (3), 374–379.

Nannicini, T., 2007. Simulation-based sensitivity analysis for matching estimators. *Stata Journal*, 7 (3), 334–350.

Niankara, I., 2011. *Essays in Risk and Applied Bayesian Econometric Modeling*, PhD thesis, Oklahoma State University. Available from: https://shareok.org/bitstream/handle/11244/6721/Department%20of%20Economics%20and%20Legal%20Studies_13.pdf?sequence=1. [Accessed 27 April 2013].

Pan, Y., Li, S. and Tse, D., 1999. The impact of order and mode of market entry on profitability and market share. *Journal of International Business Studies*, 30 (1), 81–104.

Peel, M., 1989. The going-concern qualification debate: some UK evidence. *British Accounting Review*, 21 (4), 329–350.

Peel, M. and Makepeace, G., 2012. Differential audit quality, propensity score matching and Rosenbaum bounds for confounding variables. *Journal of Business Finance & Accounting*, 39 (5&6), 606–648.

Rivers, D. and Vuong, Q., 1988. Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, 39 (3), 347–366.

Robinson, D., Robinson M. and Sisneros, C., 2012. Bankruptcy outcomes: does the board matter? *Advances in Accounting*, 28 (2), 270–278.

Rock, S., Sedo, S. and Willenborg, M., 2001. Analyst following and count-data econometrics. *Journal of Accounting and Economics*, 30 (3), 351–373.

Roodman, D., 2011. Estimating fully observed recursive mixed-process models with cmp. *Stata Journal*, 11 (2), 1–55.

Roodman, D., 2013. Cmp: Stata module to implement conditional (recursive) mixed process estimator. Available from: http:// ideas.repec.org/c/boc/bocode/s456882.html [Accessed 6 March 2013].

Rosenbaum, P., 1991. Discussing hidden bias in observational studies. *Annals of Internal Medicine*, 115 (11), 901–905.

Rosenbaum, P., 2005, Heterogeneity and causality: unit heterogeneity and design sensitivity in observational studies. *American Statistician*, 59 (2), 147–152.

Rosenbaum, P., 2010. *Design of Observational Studies*. Series in Statistics. New York: Springer.

Rosenbaum, P. and Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70 (1), 51–55.

Rubin, D. 2001. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2 (1), 169–188.

Shaw, K., 2012. CEO incentives and the cost of debt. *Review of Quantitative Finance and Accounting*, 38 (3), 323–346.

Srinidhi, B., Gul, F. and Tsui, J., 2011. Female directors and earnings quality. *Contemporary Accounting Research*, 28 (5), 1610–1644.

Stock, J. and Trebbi, F., 2003. Retrospectives: who invented instrumental variable regression? *Journal of Economic Perspectives*, 17 (3), 177–194.

Tucker, J., 2011. Selection bias and econometric remedies in accounting and finance research. *Journal of Accounting Literature*, 29, 31–57.

Vargas, J., 2012. Binding constraints: does firm size matter? Working paper. Available from: http://econpapers.repec.org/paper/pramprapa/41286.htm [Accessed 4 March 2013].

Vasquez, W., 2011. Household preferences and governance of water services. Working paper. Available from: http://www.ifpri.org/publication/household-preferences-and-governance-water-services [Accessed 23 April 2013].

Weiss, D., 2010. Cost behavior and analysts earnings forecasts. *Accounting Review*, 85 (4), 1441–1471.

Whisenant, S., Sankaraguruswamy, S. and Raghunandan, K., 2003. Evidence on the joint determination of audit and non-audit fees. *Journal of Accounting Research*, 41 (4), 721–744.

Wilde, J., 2008. A simple representation of the Bera-Jarque-Lee test for probit models. *Economics Letters*, 101(2), 119–121.

Wooldridge, J., 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. London: MIT Press.

Wu, W., Wu, C., Zhou, C. and Wu, J., 2012. Political connections, tax benefits and firm performance: evidence from China. *Journal of Accounting and Public Policy*, 31 (3), 277–300.

Zahabi, S., Miranda-Moreno, L., Patterson, Z. and Barla, P., 2012. Evaluating the effects of land use and strategies for parking and transit supply on mode choice of downtown commuters. *Journal of Transport and Land Use*, 5 (2), 103–119.

Table 1. Methods for controlling for unobserved bias.

Outcome variable	**Endogenous/selection variable type**	**Method of estimation**	**Statistical package**	**Statistical commands**
Continuous	Continuous	IV 2SLS	Stata	ivreg2
Continuous	Binary	Two-step	Stata	treatreg or etregress
Continuous	Binary	Maximum likelihood	Stata	treatreg or etregress
Continuous	Multinomial	Maximum simulated likelihood	Stata	mtreatreg
Continuous	Ordinal	Maximum likelihood	Stata	cmp
Binary	Binary	IV two-stage	Stata	ivtreatreg
Binary	Binary	IV two-stage + probit instrument	Stata	ivtreatreg
Binary	Binary	Maximum likelihood	Stata	ssm
Binary	Binary	Maximum likelihood	Stata	biprobittreat
Binary	Multinomial	Maximum simulated likelihood	Stata	mtreatreg
Binary	Ordinal	Maximum likelihood	Stata	cmp
Binary	Continuous	Two-step	Stata	ivprobit
Binary	Continuous	Maximum likelihood	Stata	ivprobit
Multinomial	Binary	Bayesian	R	endogMNP
Multinomial	Multinomial	Bayesian	R	endogMNP
Multinomial	Binary	Maximum likelihood	Stata	cmp
Multinomial	Multinomial	Maximum likelihood	Stata	cmp
Multinomial	Ordinal	Maximum likelihood	Stata	cmp
Multinomial	Continuous	Maximum likelihood	Stata	cmp
Ordinal	Binary	Maximum likelihood	Stata	ssm
Ordinal	Multinomial	Maximum likelihood	Stata	cmp
Ordinal	Ordinal	Maximum likelihood	Stata	cmp
Ordinal	Continuous	Maximum likelihood	Stata	cmp
Count	Binary	Maximum likelihood	Stata	ssm
Count	Binary	Maximum likelihood	Stata	etpoisson
Count	Multinomial	Maximum simulated likelihood	Stata	mtreatreg
Quantile	Binary	IV two-stage	Stata	ivqte
Quantile	Continuous	Two-step	Stata	cqiv

Table 2. Sensitivity methods for potential hidden bias

Panel A: methods for regression models			
Outcome variable	**Endogenous variable type**	**Statistical package**	**Statistical commands**
Continuous	Any	Excel*	formatted
Binary	Any	Stata	gsa
Continuous	Any	Stata	gsa
Panel B: methods for propensity score matching estimators			
Outcome variable	**Endogenous variable type**	**Statistical package**	**Statistical commands**
Continuous	Binary	Stata	rbounds
Binary	Binary	Stata	mhbounds
Binary	Binary	Stata	sensatt

Note:
* The formatted Excel spreadsheet is freely available (see note 23 of the current paper).