Addressing Unobserved Selection Bias in Accounting Studies: The Bias Minimization Method

Abstract

This note explains the minimum-biased estimator (MBE), which accounting researchers can use to analyse the robustness of regression or propensity score matched treatment estimates to unobserved selection (endogeneity) bias. Based on the principles of the Heckman treatment model, the MBE entails estimating matched treatment effects within a range of propensity scores that minimizes unobserved selection bias. A major advantage of the MBE is that an instrumental variable is not required. The potential utility of the MBE in accounting studies is highlighted, and a familiar empirical illustration is provided.

Keywords: unobserved selection bias, propensity score matching, bias minimization method, empirical illustration

JEL descriptors: C10, C21, M40, M41

Addressing Unobserved Selection Bias in Accounting Studies: The Bias Minimization Method

1. Introduction and Background

A central concern in observational (non-randomized) treatment effect studies is potential unobserved selection (endogeneity) bias. Such bias occurs when an omitted variable correlates with both the selection (treatment) variable and the dependent variable, leading to a biased estimate of the impact of the selection variable (treatment effect) on the outcome of interest. This note describes a statistical method whereby accounting researchers employing regression or propensity score matching (PSM) estimators can assess the robustness of treatment estimates to unobserved (hidden) selection bias. Specifically, hidden bias has its largest effect on treatment estimates for matched cases in the tails of the distribution of selection (treatment) probabilities that are predicted by a probit model, whereas bias is minimized for matched observations with selection probabilities (propensity scores) of 0.5. Consequently, to minimize the potential impact of omitted variable bias, PSM treatment estimates are confined to matched cases with selection probabilities (p) within a defined interval around 0.5 (e.g. $0.33 < p < 0.67$). Millimet and Tchernis (2013, p. 983) refer to this method as the 'minimum-biased estimator' (MBE). Standard PSM and/or regression treatment estimates can then be compared to MBE treatment effects to assess their robustness to potential hidden selection bias.

Accounting studies frequently employ regression and/or PSM methods to estimate treatment effects, with the former estimating the average treatment effect and the latter the average treatment effect on the treated. Both estimators assume that treatment estimates are not confounded by unobserved selection bias, known as the conditional independence assumption (CIA). Importantly, Millimet and Tchernis (2013) stress that an advantage of the MBE is that it provides unbiased treatment estimates when the CIA holds, but that it minimizes 'the bias associated with estimators that require the CIA when this assumption fails' (p. 983). Available alternative estimators that account for omitted variable bias rely on quasi-random variation produced by an instrumental variable.[1] For regression treatment estimates, the Heckman treatment model can be employed to control for unobserved selection bias (e.g. Leuz & Verrecchia, 2000). Although the Heckman estimator is technically identified via functional form restrictions, unlike the MBE, it requires a valid instrumental variable for credible implementation[2] (Lennox, Francis, &Wang, 2012). Such a variable is a significant determinant of the selection (treatment) variable but is not associated with the dependent variable, other than via its correlation with the treatment variable. Valid instrumental variables are usually hard (if not impossible) to find in accounting studies (Larcker & Rusticus, 2010; Lennox et al., 2012). Whilst not controlling for hidden bias, the MBE draws on both the Heckman model and PSM to estimate treatment effects where hidden selection bias is minimized.

Several sensitivity methods can be employed with regression and PSM estimators to ascertain how large an impact a simulated unobserved variable must have to render treatment effects statistically insignificant (for a review of the methods, see Peel, 2014). As with the MBE, sensitivity techniques do not require an instrumental variable; but the MBE differs in its approach, in that it aims to gauge the influence of actual (as opposed to simulated) hidden bias. Given the increasing use of PSM in accounting research (Tucker, 2010), the MBE has high potential utility. In addition to continuous, ordinal and binary dependent variables, the MBE can also be employed in studies that use PSM in combination with difference-in-difference

[1] The standard method to control for endogenous continuous variables is a two-stage-least-squares regression (Larcker & Rusticus, 2010). Unlike the Heckman model, which technically does not require an instrumental variable, the two-stage-least-squares and similar techniques (Larcker & Rusticus, 2010) do require an instrumental variable (exclusion restriction).

[2] With regard to Heckman selection models, Lennox et al. (2012) comment, 'We demonstrate empirically that the selection model is fragile and that results can be non-robust and therefore unreliable when researchers choose exclusion restrictions in an *ad hoc* fashion or choose none at all' (p. 590).

estimators (e.g. Bandick & Patrick, 2011). Note, however, that the MBE is not a panacea for the omitted variable problem and, as with all methods, is subject to limitations, not least that it may be impractical to implement in small samples.

1.1. PSM and Regression Adjustment

To inform the following exposition of the MBE, this section briefly describes the method of PSM (Rosenbaum & Rubin, 1983), that is regularly used in accounting studies (Tucker, 2010), together with regression-adjusted PSM. With PSM, no model functional form is required. Also, because estimates are confined to the common support (where treated and control subjects have similar attributes), PSM does not rely on linearity assumptions to extrapolate treatment effects outside this region. Typically, nearest-neighbour (NN) matching is employed. Using logit or probit model predicted probabilities, treated subjects are matched (with or without replacement) to untreated ones with the nearest probabilities. Closer matching may be obtained by using a caliper to specify the maximum difference in probabilities for matching treated and untreated subjects. After matching, the average treatment effect on the treated (ATT) can be estimated as a simple difference in means.

However, a potential limitation of NN matching (which by design is amplified when using the MBE) is that small samples may result, thereby reducing the power of statistical tests. This problem may be at least partly circumvented by employing weighted PSM methods that use all observations in the common support by weighting untreated outcomes according to the distance between the propensity scores of treated and untreated cases. Inverse probability weighting is the simplest method, in that it employs inverse selection probabilities to weight outcomes in calculating the ATT (Curtis, Hammill, Eisenstein, Kramer, & Anstrom, 2007). A more complex method involves applying kernel-function estimated weights (Heckman, Ichimura, & Todd, 1998). All PSM methods can be employed with the MBE (Black & Smith, 2004; Eren, 2007; Millimet & Tchernis, 2013). As explained herein, NN matching with a caliper is preferable, because the closer the treated and untreated subjects are matched, the greater is the potential bias reduction using the MBE.

In accounting studies, the standard approach is to estimate a regression model in the NN-matched sample (e.g. Ittonen, Johnstone, & Myllymäki, 2015; Minutti-Meza, 2013) to account for any covariate imbalance post matching (Imbens & Wooldridge, 2009, p. 37). A less well-known rationale for this procedure is that treatment estimates are doubly robust 'in the sense that … if either the matching or the parametric model is correct, but not necessarily both, causal estimates will still be consistent' (Ho, Kosuke, King, & Stuart, 2007, p. 215). Hence, when employing PSM within a specified radius (herein), a further MBE robustness test is to estimate a standard regression model in the matched sample.

1.2. The Potential Utility of the MBE in Accounting Research

Given the increasing employment of PSM in accounting studies (Tucker, 2010), the MBE is a useful robustness method, in that it enables assessment of the potential influence of hidden bias on PSM treatment effects. Furthermore, a typical research approach in accounting studies is to compare ordinary least squares (OLS) regression average treatment estimates to matched ATT ones to assess if treatment inferences are consistent (e.g. Lennox & Pittman, 2010). For instance, Ittonen et al. (2015), in examining the relationship between audit partner public-client specialization and abnormal accruals, report that their PSM results support their regression findings. Importantly, they state that the Heckman approach was not adopted because an instrumental variable was unavailable (Ittonen et al., 2015, p. 629).[3] In such studies, the MBE can be used to gauge the potential impact of hidden selection bias.

[3] Lennox and Pittman (2010, p. 236) and Minutti-Meza (2013, p. 793) make similar points.

The application of the MBE may also be particularly informative when regression and PSM treatment estimates provide conflicting evidence. Specifically, other things equal, differences between regression and PSM treatment estimates should arise because of observed bias, which is attributable to the unmatched characteristics[4] of treated and untreated observations in the regression model (Lawrence, Minutti-Meza, & Zhang, 2011, p. 268) and which PSM controls for. In these cases, the MBE can be employed to assess the robustness of PSM treatment effects to hidden selection bias. For example, in contrast to studies that report a positive association between CEO equity incentives and accounting irregularities, Armstrong, Jagolinzer, and Larcker (2010) use PSM to find evidence of a negative relationship. Additionally, contrary to prior research using standard regression models, two studies use PSM and report no significant association between auditor industry specialization (Minutti-Meza, 2013), or big 4 audits (Lawrence, et al., 2011), on audit quality outcomes.

2. The Bias Minimization Method

2.1. Linking PSM and the Heckman Treatment Model

Both the Heckman treatment model and PSM employ a selection into treatment regression model. Given that with a selection model, the actual outcomes minus the predicted ones equal the residuals (the unexplained variation in the dependent variable), the Heckman estimator uses the generalized residuals (inverse Mills ratios) from the first step probit selection model as a surrogate for unobserved variables in the second step OLS outcome model (e.g. Lennox et al., 2012; Tucker, 2010; Peel, 2014). The Heckman method is predicated on the errors of the probit and OLS models being jointly and normally distributed. Probit models employ the cumulative distribution function of the standard normal distribution, and model errors are assumed to follow the standard normal distribution.

[4] For instance, this may occur when the linearity assumption does not hold outside the common support region.

As previously discussed, PSM uses the predicted probabilities from a selection model to match treated and untreated subjects. Importantly, in the case of the MBE, where PSM is based on a probit selection model (as per the Heckman approach), a matched case has a predicted probability and an associated residual (inverse Mills ratio), which is a proxy for omitted variables. As described herein, the rationale of the MBE is that the impact of omitted variable bias (probit model residuals) is minimized for propensity score matched subjects over a specified range of probabilities centered on 0.5.

2.2. Derivation of the Bias Minimization Method

Rosenbaum (2005) demonstrates analytically that matching may alleviate the impact of unobserved bias in 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’ (p. 6). As shown by Black and Smith (2004, pp. 111–113) and Millimet and Tchernis (2013, pp. 986–989), unobserved selection bias can be minimized by matching treated and untreated subjects with predicted probabilities (from a probit selection into treatment model), which are within specified intervals around 0.5. Given normally distributed errors, the unobserved selection bias associated with a given probability B(*PX*) for matched ATT treatment estimates is:

$$\mathrm{B}\big(P(X)\big) = \rho\sigma_{\varepsilon_0}\frac{\phi\big(h(X)\big)}{\Phi\big(h(X)\big)\big[1-\Phi\big(h(X)\big)\big]} \tag{1}$$

where X is a vector of matching covariates used in a probit treatment selection model, $h(X)$ are predicted selection model values (not transformed into probabilities) using the coefficients estimated with the vector of X variables, $P(X)$ are the predicted probabilities, ϕ and Φ are the normal density and cumulative distribution functions and ρ, and σ estimates the covariance (correlation) between the ratio and the error of the outcome model (ε_0). With reference to Black and Smith (2004) and Heckman and Navarro-Lozano (2004), Millimet and Tchernis (2013, pp. 987–988) give a full derivation for (1). From (1), any unobserved selection bias is minimized when $P = 0.5$ or equivalently when $h(X) = 0$. As stressed by Black and Smith (2004, p. 111) it 'will have its largest effects on bias for values of the propensity score in the tails of the distribution.'[5]

Because potential bias reduction is most marked in intervals centered on 0.5, Black and Smith (2004) suggest that matching estimators should be applied to the sample with selection probabilities within the range of 0.33 to 0.67. However, if sufficient observations are available, a narrower radius would provide an even stricter test, because bias reduces as probabilities converge on 0.5. As discussed herein, the downside is that fewer matched observations occur as the radius narrows. As Millimet and Tchernis (2013) note, finer intervals 'should reduce the bias at the expense of higher variance' (p. 989). Importantly, Black and Smith (2004) emphasize that whilst the MBE assumes joint normality of error terms, the intuition of the method 'does not depend on distributions' (of errors), in that 'when the probability of being in the treatment group is high, unobservable factors on average play a larger role than for probabilities near 0.5. Thus, when matching estimators must rely on the right tail of the distribution of propensity scores in the comparison group, the selection bias may be considerable' (p. 113). A similar rationale applies to the left tail of the distribution. Millimet and Tchernis (2013) make an even stronger case that joint normality is not required to establish that the ATT bias-minimizing propensity score (referred to as BMPS, P^*) is 0.5. They emphasize the following:

> it is trivial to show that $P^* = 0.5$ for the ATT … under a wider class of models than joint normality… Because the bias of the ATT … is minimized by minimizing the bias

[5] At one level, this is similar to the removal of data outliers which have a disproportionate influence on regression estimates (Leone, Minutti-Meza, & Wasley, 2015) However, the MBE differs in that unobserved bias, which has its largest effect in the tails of the probability distribution, does not arise from data outliers, but from the distribution of residuals.

> for each component obtaining draws from a particular trivariate normal distribution *and* the BMPS is one-half within each component, the bias of the ATT … is minimized at $P^* = 0.5$. Furthermore, because a mixture of a sufficient number of trivariate normal distributions can approximate almost any joint distribution, this implies that joint normality is not needed to conclude that one-half is the BMPS for the ATT. (p. 990)

Hence, the MBE is a viable technique to assess the robustness of PSM treatment effects to hidden selection bias. Other things equal, relative to significant treatment effects estimated with standard methods, the interpretation of MBE treatment estimates is as follows: (a) MBE treatment effects are similar and significant, implying unobserved bias is not a significant threat and offering support for the CIA; (b) MBE treatment effects are higher and significant, implying that standard treatment estimates are subject to downward bias (under-estimated) due to the presence of an unobserved correlated variable[6]; (c) MBE treatment effects are lower and significant, implying that standard treatment estimates are subject to upward bias (over-estimated) due to the presence of an unobserved correlated variable[7]; and (d) MBE treatment effects are statistically insignificant and close to zero, implying that standard significant treatment estimates arise as a consequence of omitted variable bias. For instance, a statistically significant standard PSM treatment effect (e.g. a big 4 audit fee premium) of 10% and significant MBE treatment effects of 5% (15%) are consistent with the standard PSM treatment estimates being upwardly (downwardly) biased due to an unobserved correlated variable.

2.3. Potential Limitations

Unlike the MBE, the Heckman treatment model controls directly for unobserved bias if implemented correctly and thereby provides a bias-free estimate of the treatment effect.

[6] Specifically, by omitting a variable that is positively (negatively) correlated with the dependent variable and negatively (positively) correlated with the treatment variable, the standard treatment estimate is subject to downward bias. Hence, including such a variable increases (inflates) the treatment estimate.
[7] Specifically, by omitting a variable that is positively (negatively) correlated with the dependent variable and positively (negatively) correlated with the treatment variable, the standard treatment estimate is subject to upward bias. Hence, including such a variable reduces (deflates) the treatment estimate.

Furthermore, an axiomatic consequence of employing PSM is that treatment estimates are based on smaller samples, such that the power of statistical tests is reduced (e.g. Lawrence et al., 2011), although they are generally valid unless very small samples are employed. This may lead, for example, to an insignificant PSM treatment effect that might be significant in a larger sample with all other things equal. Inherently, the MBE exacerbates this problem, such that in practice it may be impracticable to implement the method with confidence in small matched samples. However, in addition to using NN matching with replacement, the sample size may be increased substantially by employing PSM inverse probability or kernel-weighting estimators (Black & Smith, 2004; Millimet & Tchernis, 2013). For instance, when investigating the union wage effect, Eren (2007, p. 776) reports statistically significant treatment effects of 0.286 (0.271) using NN (kernel) PSM estimators, with a kernel sample size (3,550) that is 3.8 times larger than the NN one (937). For small-sample treatment estimates, bootstrap standard errors may also be used as a robustness test (Minutti-Meza, 2013, p. 812).

By construction, a more general limitation of PSM and hence the MBE is that information on observations outside the common support (which *per se* may be of interest[8]) are discarded. This feature also underpins the difference between the average treatment effect and the ATT. A further limitation is that, whilst standard PSM ATT estimates are representative of ATT population means, MBE ATT estimates are constrained to a range of propensity scores in which bias is minimized. Hence, they are representative only of the specific sample (population). A further potential limitation is that the difference between standard PSM and MBE treatment

[8] For instance, in 2014 only two FTSE 100 companies had non-big 4 auditors. Hence in a PSM study of quoted companies, with only two counterfactuals, big 4 FTSE 100 auditees would be (at least largely) excluded; notwithstanding that they may be of particular interest. For regression models, FTSE 100 treatment (big 4) estimates would rely on the linearity assumption.

estimates may at least partly reflect heterogeneity.[9] Heterogeneity occurs when treatment effects vary in accord with subject-specific (idiosyncratic) differential responses[10] to selection into treatment (within-group variation). However, even in this case, unobserved bias should still be minimized using the MBE, because it identifies the parameter (ATT) that 'can be estimated with the least bias' (Millimet and Tchernis, 2013, p. 988).

Tables 1-3 about here

2.4. Empirical Illustration

This section provides an empirical illustration of the MBE with respect to the estimation of big 4 audit fee treatment effects. An advantage of using audit fee data is that there is reasonable consensus on the control variables that should be included in the model and about the expected relationship between explanatory variables and audit fees. Corporate size is (unsurprisingly) the principal determinant of both selection into treatment and audit fees (e.g. Minutti-Meza, 2013, p. 789). In addition, a large number of observations (5,640) are available to estimate treatment effects. Following typical accounting treatment effect studies, as previously mentioned, big 4 premium regression estimates are first compared to standard PSM and PSM regression-adjusted ones. The robustness of these treatment estimates to potential unobserved selection bias is then assessed with reference to MBE treatment estimates. The data are drawn from those used in Peel

[9] In this context, it is possible that subjects which are comparatively indifferent regarding selection into treatment (i.e. p = 0.5) are those where treatment is less important; and hence ATT estimates for these cases may not represent population ones. However, *a priori*, there is no reason to suspect that treated observations with selection probabilities of 0.5 will not attract the same treatment effect (e.g. a big 4 premium) as other treated observations. Note also that the rationale of the MBE is that the estimated ATT is the one subject to the least unobserved selection bias. As emphasised by Millimet and Tchernis (2013), where 'relevant regressors are omitted from the model, the MB estimator improves upon the performance of commonly used estimators that require the CIA' (p. 983).

[10] An example is provided Daske, Hail, Leuz, and Verdi (2013), who state that, when studying the impact of the adoption of International Accounting Standards (IAS) and International Financial Reporting Standards (IFRS), 'some firms may make very few changes and adopt IAS/IFRS more in name, while for others the change in standards could be part of a strategy to increase their commitment to transparency' (p. 495). Interestingly, Athey and Imbens (2015) specify a number of machine learning 'data-driven' tree-based methods which aim to detect/quantify heterogeneous average treatment effects over sub-populations (leaves) of the population. Note also that unobserved subject-specific heterogeneity is also a form of omitted variable (endogeneity) bias and that panel (including difference-in-difference) estimators can account for time invariant unobserved heterogeneity (Wooldridge, 2010).

(2013) and comprise 5,640 UK private independent non-financial companies that had year-ends in 2005/6, that were live (not failed/dissolved/dormant) and that had not switched auditors. Of the 5,640 companies, 2,890 are big 4 auditees, and the remainder were audited by the four largest mid-tier (mid 4) auditors. The OLS model is:

$$\text{LnFEE} = \alpha_0 + \beta_1\text{LnSAL} + \beta_2\text{LnTA} + \beta_3\text{SQSUBS} + \beta_4\text{EXPSAL} + \beta_5\text{QUAL} + \beta_6\text{PBAL} + \beta_7\text{CONLIAB} + \beta_8\text{EXITEM} + \beta_9\text{RSAL} + \beta_{10}\text{LOSS} + \beta_{11}\text{TLTA} + \beta_{12}\text{LOND} + \beta_{13}\text{BUSY} + \beta_{14}\text{YR} + \beta_{15}\text{IND} + \beta_{16}\text{BIG4} + \varepsilon \qquad (2)$$

where LnFEE is the natural log of audit fees, LnSAL is the natural log of sales, LnTA is the natural log of total assets, SQSUBS is the square root of the number of subsidiaries, EXPSAL is the ratio of non-UK to total turnover, RSAL is the ratio of profit before tax to sales, and TLTA is the ratio of total liabilities to total assets. Binary variables are QUAL, which equals 1 if the accounts received an audit qualification; PBAL, which equals 1 if the accounts disclosed a post-balance sheet event; CONLIAB, which equals 1 if the accounts disclosed contingent liabilities; EXITEM, which equals 1 if the accounts disclosed exceptional or extraordinary items; LOND, which equals 1 if the company is located in London; BUSY, which equals 1 if the company's account year-end is in December or March; YR, which equals 1 if the company's account year-end is in 2006 (2005 coded as zero); LOSS which equals 1 if the company is loss-making; and BIG4, which equals 1 if the company has a big 4 auditor. As shown in Table 1, IND represents industry dummies. Binary variables are coded zero for companies without the attribute.

Table 1 defines the explanatory variables and reports parameters for the OLS LnFEE model and the BIG4 probit selection model. NN matching without replacement, imposing a fine caliper (0.001), is employed.[11] Although the big 4 auditee means of LnSAL and LnTA are significantly larger than those of their mid 4 counterparts (see Table 3), the probit estimates show

[11] Following Rubin (1997, p. 761), all variables are included in the OLS and probit models, whether or not they are significant.

that LnSAL has a small but significant negative impact on big 4 auditor selection after controlling for LnTA, which hinders close matching of both size variables. Hence, following the normal approach of accounting studies that employ PSM, the previously described regression-adjusted PSM is applied. The OLS model specified in Table 1 is re-estimated in each NN matched sample. Table 2 reports big 4 mean premium estimates,[12] with results labelled for each model (M) employed. The first three rows in the table show standard regression, PSM and regression-adjusted PSM premium estimates. The standard regression (M1a) big 4 premium estimate is highly significant at 11.11%, but the standard PSM one (M1b) falls to 3.66% and is statistically insignificant.

However, significant size imbalance remains after matching, with both LnTA and LnSAL being larger for mid 4 auditees, thus downwardly biasing the premium estimate. When the OLS regression model specified in Table 1 is applied to the matched sample (M1c), the premium rises to 9.92% and is highly significant ($p < 0.001$). As Table 2 reveals, a similar picture emerges when MBE estimates are confined to the interval of 0.33, 0.67 as per Black and Smith (2004), with a size imbalance again remaining after matching. The MBE PSM premium (3.98%) is statistically insignificant (M2a), but the premium estimate of its regression-adjusted MBE counterpart (9.24%) is highly significant (M2b). Table 3 reports covariate balance statistics for unmatched big 4 and mid 4 auditees and matched ones for Model 3a. It shows that the application of a finer radius (0.4, 0.6) resulted in close matching[13] on the size variables. The mean values of LnSAL and LnTA are similar for big 4 and mid 4 auditees and do not differ significantly. As Table 2 shows, the MBE PSM premium estimate of 9.34% (M3a) is now statistically significant ($p = 0.029$), with the MBE PSM regression-adjusted one (M3b) being

[12] Please see the online Supplement for all unreported results, including covariate balance statistics.
[13] As with M1b and M2a (other than for size), the only matching variable to differ significantly for M3a is NID.

highly significant and marginally higher (10.42%). In summary, as described in Section 2.2, the similarity of the standard PSM regression-adjusted premium estimates and the MBE ones support the CIA, implying that unobserved bias does not impact adversely on ATT treatment effects.[14] In particular, the MBE premiums estimated within the finest interval (0.4, 0.6), where there is close covariate balance, provide especially strong support for this inference.

3. Conclusion

Non-randomized (observational) treatment effect studies are prone to unobserved selection bias. The MBE provides accounting researchers with a viable technique to assess the robustness of conventionally estimated PSM treatment effects to potential unobserved selection bias. Unlike the Heckman treatment model, which is predicated on the assumption of joint normality of errors and for which an appropriate instrumental variable is required for credible implementation, the application of the MBE is not subject to these constraints. However, the MBE is not a panacea for the omitted variable problem. By design, it leads to smaller samples than does standard PSM, such that it may be impractical to implement with confidence in some applications. In practice, this problem may be at least partially alleviated by employing weighted PSM approaches, such as kernel and inverse probability weighting estimators.[15]

Supplemental Material

Supplemental material for this article can be accessed on the Taylor & Francis website.

[14] As a stability test, the OLS models were re-estimated with the Stata robust regression M-estimator (*rreg*), which is resistant to outliers. Though marginally larger than for their OLS counterparts, all BIG4 coefficients are highly significant ($p < 0.001$), with the associated premiums exhibiting the same relative size rankings as the OLS ones reported in Table 2: 11.88%, 10.99%, 9.98%, and 11.32% for models 1a, 1c, 2b, and 3b, respectively (see Supplement). Similar results were obtained when employing the alternative user-written Stata MM robust estimator (*mmreg*) of Verardi and Croux (2009).

[15] In this context, McCarthy, Millimet, and Tchernis (2014) have written a Stata programme, *bmte* (bias-minimizing treatment effects), which *inter alia* employs a PSM inverse probability-weighting ATT estimator with the MBE.

References

Armstrong, C., Jagolinzer, A., & Larcker, D. (2010). Chief executive officer equity incentives and accounting irregularities. *Journal of Accounting Research*, *48*(2), 225–271.

Athey, S., & Imbens, G. (2015). *Recursive partitioning for heterogeneous causal effects*. Unpublished manuscript. Retrieved March 28, 2016, from http://arxiv.org/pdf/1504.01132.pdf

Bandick, R., & Patrick, K. (2011). Employment effects of foreign acquisition. *International Review of Economics and Finance*, *20*(2), 211–224.

Black, D., & Smith, J. (2004). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics*, *121*(1&2), 99–124.

Curtis, L., Hammill, B., Eisenstein, E., Kramer, J., & Anstrom, K. (2007). Using inverse probability-weighted estimators in comparative effectiveness analyses with observational databases. *Medical Care*, *45*(10), 103–107.

Daske, H., Hail, L., Leuz, C., & Verdi, R. (2013). Adopting a label: Heterogeneity in the economic consequences around IAS/IFRS adoptions. *Journal of Accounting Research*, *51*(3), 495–547.

Eren, O. (2007). Measuring the union/non-union wage gap using propensity score matching. *Industrial Relations: A Journal of Economy and Society*, *46*(4), 766–780.

Heckman, J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies*, *65*(2), 2612–2694.

Heckman, J., & Navarro-Lozano, S. (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics, 86*(1), 30–57.

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

Imbens, G., & Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86.

Ittonen, K., Johnstone, K., & Myllymäki, E-R. (2015). Audit partner public-client specialisation and client abnormal accruals. *European Accounting Review*, *24*(3), 607–633.

Larcker, D., & 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., & Zhang, P. (2011). Can big 4 versus non-big 4 differences in audit quality proxies be attributed to client characteristics? *The Accounting Review*, *86*(1), 259–286.

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

Lennox, C., & Pittman, J. (2010). Big five audits and accounting fraud. *Contemporary Accounting Research*, *27*(1), 209–247.

Leone, A., Minutti-Meza, M., & Wasley, C. (2015). *Influential observations and inference in accounting research*. Simon Business School working paper. Retrieved April 2, 2016, from http://ssrn.com/abstract=2407967

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

McCarthy, I., Millimet, D., & Tchernis, R. (2014). The bmte command: Methods for the estimation of treatment effects when exclusion restrictions are unavailable. *The Stata Journal*, *14*(3), 670–683.

Millimet, D., & Tchernis, R. (2013). Estimation of treatment effects without an exclusion restriction: With an application to the analysis of the school breakfast program. *Journal of Applied Econometrics*, *28*(6), 982–1017.

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

Peel, M. (2013). The pricing of initial audit engagements by big 4 and leading mid-tier auditors. *Accounting and Business Research*, *43*(6), 636–659.

Peel, M. (2014). Addressing unobserved endogeneity bias in accounting studies: Control and sensitivity methods by variable type. *Accounting and Business Research*, *44*(5), 545–571.

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

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

Rubin, D. (1997). Estimating causal effects from large data sets using the propensity score. *Annals of Internal Medicine*, *8*(2), 757–763.

Tucker, J. (2010). Selection bias and econometric remedies in accounting and finance research. *Journal of Accounting Literature*, *29*, 31–57.

Verardi, V., & Croux, C. (2009). Robust regression in Stata. *The Stata Journal*, *9*(3), 439–453.

Wooldridge, J. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge, MA: MIT Press.

Table 1. OLS audit fee and big 4 probit regression model coefficients

Model 1: Dependent Variable = LnFEE Model 2: Dependent Variable = BIG4		
Variable definitions	(1) OLS (n = 5,640)	(2) Probit (n = 5,640)
LnSAL: natural log sales (£)	0.265**	-0.038**
LnTA: natural log total assets (£)	0.124**	0.123**
SQSUBS: square root of number of subsidiaries	0.170**	0.052**
EXPSAL: ratio of non-UK to total turnover	0.524**	0.337**
QUAL: 1 if audit qualification‡	0.102**	-0.154*
PBAL: 1 if disclosed post-balance sheet event‡	0.182**	0.129*
CONLIAB: 1 if disclosed contingent liabilities‡	0.082**	-0.081
EXITEM: 1 if disclosed exceptional/extraordinary items‡	0.066**	0.001
RSAL: ratio of profit before tax to sales	-0.4E-3**	0.3E-3
LOSS: 1 if loss-making‡	0.182**	0.200**
TLTA: ratio of total liabilities to total assets	0.011**	0.014**
LOND: 1 if company is located in London‡	0.313**	-0.162**
BUSY: 1 if year-end is in December or March‡	0.013	0.080*
YR: 1 if year-end is 2006‡	-0.036*	-0.074*
MAN: 1 if manufacturing sector‡	0.036	-0.162**
RET: 1 if retail/wholesale sector‡	-0.057*	-0.215**
OIN: 1 if other industrial sector‡	-0.019	0.106
NIC: 1 if no industry code disclosed‡	0.076	0.264
SERV: 1 if service sector‡ (base case)	-	-
BIG4: 1 if big 4 auditor‡	0.105**	-
Constant	2.910**	-1.330**
R^2 /chi^2	0.760	321.4**

Notes: This table reports the ordinary least squares (OLS) and probit selection model regression coefficients for the whole sample. The OLS dependent variable is LnFEE, the natural log of audit fees. The probit model dependent variable is BIG4, coded as unity for big 4 auditors and zero otherwise.

‡ Indicates binary variables, where 0 is coded for the remaining observations.

* Indicates coefficient is statistically significant at the 5% level (two-tailed test).

** Indicates coefficient is statistically significant at the 1% level (two-tailed test).

Table 2. Propensity score matched (PSM) and regression model big 4 premium estimates

Model	Sample and method	Big 4 premium estimate	Significance of big 4 premium (p-values)	Number of companies	OLS R^2
1a	OLS all companies	11.11%**	0.000	5,640	0.760
1b	PSM all companies	3.66%	0.349	3,018	-
1c	OLS PSM all companies	9.92%**	0.000	3,018	0.700
2a	PSM 0.33 < p < 0.67	3.98%	0.309	2,898	-
2b	PSM OLS 0.33 < p < 0.67	9.24%**	0.000	2,898	0.695
3a	PSM 0.4 < p < 0.6	9.34%*	0.029	2,432	-
3b	PSM OLS 0.4 < p < 0.6	10.42%**	0.000	2,432	0.689

Notes: This table reports big 4 premiums estimates and summary statistics for the specified models (see also, the online Supplement). OLS refers to the ordinary least squares regression model specification reported in Table 1. Table 1 defines the variables. PSM refers to propensity score matching using the nearest-neighbour (NN) method without replacement and employing a caliper of 0.001. Models 2a to 3b are minimum biased estimates, where NN matching is constrained to the sub-samples, as determined by the reported selection probabilities (p). Probabilities (propensity scores) used for matching are those estimated with the probit selection model, reported in Table 1, for the whole sample (n=5,640). Sub-samples specified with reference to p are the probabilities estimated with the probit selection model, reported in Table 1, for the whole sample. The transformation e^x-1 is employed to calculate premiums, where x = the mean difference in the natural log of audit fees (LnFEE) or the OLS BIG4 regression coefficient specified to seven decimal places.

Table 3. Covariate balance (means)

	Unmatched (whole sample)			Matched propensity (p) scores: 0.4 < p < 0.6		
Variables‡	Big 4 = 1 (n = 2,890)	Big 4 = 0 (n = 2,750)	Mean difference (p-value)	Big 4 = 1 (n = 1,216)	Big 4 = 0 (n = 1,216)	Mean difference (p-value)
LnSAL	15.550	15.128	0.000***	15.245	15.267	0.798
LnTA	15.794	15.096	0.000***	15.192	15.230	0.549
SQSUBS	1.405	0.996	0.000***	0.879	0.895	0.686
EXPSAL	0.077	0.054	0.000***	0.043	0.044	0.931
QUAL	0.047	0.052	0.427	0.053	0.054	0.857
PBAL	0.152	0.102	0.000***	0.077	0.072	0.644
CONLIAB	0.252	0.226	0.026**	0.244	0.219	0.136
EXITEM	0.191	0.155	0.000***	0.142	0.139	0.816
RSAL	-1.403	-2.881	0.512	-2.157	-0.930	0.391
LOSS	0.335	0.259	0.000***	0.235	0.216	0.265
TLTA	1.054	0.954	0.313	0.872	0.827	0.529
LOND	0.244	0.271	0.017**	0.231	0.238	0.702
BUSY	0.590	0.549	0.002***	0.558	0.550	0.714
YR	0.448	0.489	0.002***	0.482	0.492	0.627
SER	0.610	0.564	0.000***	0.527	0.551	0.238
MAN	0.212	0.227	0.180	0.265	0.257	0.644
RET	0.138	0.183	0.000***	0.188	0.179	0.600
OIN	0.025	0.018	0.066*	0.014	0.013	0.861
NIC	0.014	0.008	0.039**	0.007	0.000	0.005***

‡ Table 1 defines the variables.
* Indicates statistically significant at the 10% level (two-tailed test).
** Indicates statistically significant at the 5% level (two-tailed test).
*** Indicates statistically significant at the 1% level (two-tailed test).