This article was downloaded by: [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi]
On: 29 September 2013, At: 20:11
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Accounting and Business Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rabr20

Selection bias and the Big Four premium: New evidence using Heckman and matching models

Mark A. Clatworthy [a], Gerald H. Makepeace [b] & Michael J. Peel [b]
[a] Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK Phone: +44(0)2920 874000 E-mail:
[b] Cardiff Business School
Published online: 04 Jan 2011.

To cite this article: Mark A. Clatworthy , Gerald H. Makepeace & Michael J. Peel (2009) Selection bias and the Big Four premium: New evidence using Heckman and matching models, Accounting and Business Research, 39:2, 139-166, DOI: 10.1080/00014788.2009.9663354

To link to this article: http://dx.doi.org/10.1080/00014788.2009.9663354

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Selection bias and the Big Four premium: new evidence using Heckman and matching models

Mark A. Clatworthy, Gerald H. Makepeace and Michael J. Peel*

Abstract—Many prior studies have found that large auditors charge significantly higher fees for statutory audit services, potentially resulting from higher audit quality and/or a lack of competition in the audit market. However, recent research using a Heckman two-step procedure attributes the large auditor premium to auditor selection bias. In this paper we examine the limitations of the Heckman model and estimate the large auditor (Big Four) premium using decomposition and matching methods on a large sample of UK private companies. Our analysis suggests that Heckman two-step estimates are highly sensitive to changes in sample and model specification, particularly the presence of a valid identifying variable. In contrast, the propensity score and portfolio matching methods we employ point to a persistent large auditor premium, consistent with the majority of previous studies. Conclusions of the premium vanishing when selection bias is controlled for therefore appear premature. Since the Heckman model is increasingly used in auditing and other areas of accounting research, our discussion and findings are likely to be of more general interest.

Keywords: audit fees; Big Four premium; matching estimators; propensity score matching; selection bias; two-step models

1. Introduction and background

In recent years, the competitiveness of the market for audit services has been the subject of considerable attention from the accounting profession, regulators and academic researchers. Among the main issues of concern is whether Big Four auditors command a premium when setting fees for statutory corporate audit services and, if so, whether the premium is symptomatic of a lack of competition in the audit market, or results from a higher quality product in competitive markets. In the UK, the (then) Department for Trade and Industry and the Financial Reporting Council (Oxera, 2006) estimated the Big Four premium at 18%; they concluded that it results from higher concentration and while auditor reputation is important to companies, some large UK firms have no effective choice of auditor due to significant barriers to entry. Furthermore, since the seminal contribution of Simunic (1980), a large number of studies, from a variety of markets and countries, find a premium using ordinary least squares (OLS) regression for large (Big Eight, Big Six and Big Four) auditors, for companies of various sizes (e.g. Pong and Whittington, 1994; Seetharaman et al., 2002; McMeeking et al., 2007; Clatworthy and Peel, 2007). A survey of the international empirical evidence (Moizer, 1997: 61) reports that 'the results point to a top tier fee premium of between 16 to 37%'; meanwhile, in a meta-analysis of 147 published audit fee studies, Hay et al. (2006: 176) find that 'the results on audit quality strongly support the observation that the Big 8/6/5/4 is associated with higher fees.'

Notwithstanding these findings, recent research has turned its attention to the important issue of the non-random selection of auditors and its impact on observed large auditor premiums. Some studies report that the premium paid to large auditors is larger than implied by OLS estimates when the Heckman model is employed to control for selection bias. In a study of UK listed companies, Ireland and Lennox (2002: 89) conclude that 'the large audit fee premium is more than twice as large when one controls for selection bias (53.4% compared to 19.2%)'. Moreover, based on a sample of UK companies over the period 1985 to 2002, McMeeking et al. (2006) find that the large auditor premium increases when selection bias is controlled for. Not all research using the correction for selection bias has produced the same results, however. In a recent examination of the pricing of audit services in UK local authorities, Giroux and

*Mark Clatworthy, Gerald Makepeace and Michael Peel are at Cardiff Business School. They are grateful for helpful comments provided by two anonymous reviewers, Roy Chandler, Mahmoud Ezzamel and Wally Smieliauskas and by seminar participants at Cardiff University, City University, the University of Edinburgh, the University of Exeter, the University of Nottingham, the University of Strathclyde, the 2007 ESRC Seminar in Accounting, Finance and Economics in London and the 2007 EAA Annual Congress in Lisbon. All remaining errors are the authors' own.

Correspondence should be addressed to: Mark Clatworthy, Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK. Tel.: +44(0)2920 874000. E-mail: ClatworthyMA@cardiff.ac.uk.

This paper was accepted for publication in December 2008.

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Jones (2007) report that the selection bias correction makes little difference to inferences from OLS regressions. Furthermore, in a study of UK private companies, Chaney et al. (2004) fail to find a large auditor premium after they control for potential self-selection. Using OLS regression, they find a significant positive coefficient on a large auditor (Big Five) binary variable, but no premium when a two-step Heckman procedure to control for potential self-selection of auditors is used. Indeed, they conclude (2004: 67) that 'if big 5 auditees had chosen non-big 5 auditors, their audit fees would have been higher.' Similar findings were reported by the same authors for a sample of US listed firms (Chaney et al., 2005). These findings represent a very important development in the literature since they imply that many previous studies may have erroneously reported large auditor premiums – despite recent evidence (e.g. Blokdijk et al., 2006) that the Big Four provide higher quality audits.

The purpose of this paper is to present new evidence on the Big Four auditor premium and the effects of auditor selection for a large sample (36,674) of independent private UK firms by employing two-stage estimators and new decomposition and matching methods. Although the Heckman estimator represents an innovative correction for selection bias, it requires a valid additional identifying variable (instrument) for reliable implementation of the method; however, such variables are often extremely difficult to obtain in practice.[1] Since previous studies using the Heckman estimator do not focus on this important issue, this paper reports the sensitivity of estimates to the inclusion/exclusion of such an instrument. Furthermore, the Heckman estimator has been employed uncritically in the accounting literature to date, but analyses in other social science research suggest that it is highly sensitive to model specification, in contrast to OLS single-stage estimates (e.g. Hartman, 1991; Stolzenberg and Relles, 1997). This is particularly important because the Heckman model is increasingly used as a 'robustness test' for selection bias in accounting research in general and in the auditing literature in particular. If the model is not properly identified, this may lead to Heckman results lacking robustness due to severe collinearity problems (Little and Rubin, 1987; Puhani, 2000). Our analysis therefore focuses on these issues by illustrating the effects and sensitivity of the Heckman model to specification and sample variations and to the valid identification of the two-stage equations. An independent contemporaneous working paper by Francis and Lennox (2008) also examines the sensitivity of the Heckman model in estimating the large auditor premium for UK private firms. Although there is naturally some overlap between the two studies (and their results are of relevance to the current paper, so are discussed where appropriate), there are some important differences. In contrast to the present paper, the major aim of Francis and Lennox (2008) is to replicate the findings of Chaney et al. (2004) for a more recent time period, using the same models and sampling design as Chaney et al. (2004). The present paper does not aim to reproduce the results of Chaney et al. (2004), since it is based on a much larger cross sectional (rather than panel-based) sample of UK private firms and our models include additional important variables in order to obtain a more reliable and representative estimate of the Big Four premium. Importantly, our analysis also involves decomposition and matching methods not employed by Francis and Lennox (2008).

A further objective of the paper is to examine the significant methodological issue of how to estimate causal treatment effects in accounting and auditing research (the appointment of a Big Four auditor in our case) using OLS, selection and matching models. We conduct a comprehensive analysis of fee differences between Big Four and non-Big Four auditees using maximum likelihood (ML) and formal decomposition measures for selection effects not previously employed in the accounting and auditing literature. An increasingly popular approach adopted in the applied econometrics and finance literature involves matching procedures, particularly propensity score matching methods (e.g. Black and Smith, 2004; Simonsen and Skipper, 2006; Li and Prabhala, 2007), although these have yet to be employed in auditing research. Using these methods, we present new evidence on the large auditor premium using closely matched samples (in respect of size, risk and complexity) of companies audited by Big Four and non-Big Four auditors. In summary, our paper extends extant research by identifying potential weaknesses of the Heckman model and presenting alternative procedures to estimate the Big Four premium.

Our work contributes to the empirical literature on selection bias and the Big Four premium by using the largest sample of UK firms yet studied. The richness of our data set and large sample size allow us to subject this important issue to considerable scrutiny.[2] Our results suggest that two-step

[1] For instance, Neumayer (2003: 655) notes 'The problem is that such an exclusionary variable is frequently impossible to find'; moreover, Bryson et al. (2002: 9) state that 'the identification of a suitable instrument is often a significant practical obstacle to successful implementation.'

[2] At 36,674 observations, our sample is substantially larger than the largest (6,198 observations) in the meta-analysis of over 140 audit fee studies by Hay et al. (2006). Moreover, our dataset includes a more comprehensive set of variables than prior studies of selection bias and private UK company audit fees. In particular, the model reported by Chaney et al. (2004) excludes the number of subsidiaries and a second corporate size variable (sales), both of which have been found important in previous research.

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

corrections for selection bias in audit fee models are very sensitive to model specification (including the absence of an identifying variable), and to the sample used – findings consistent with empirical results in applications of two-step models in other fields (e.g. Winship and Mare, 1992; Stolzenberg and Relles, 1997; Leung and Yu, 2000). Using more consistent propensity score and pre-processed portfolio matching approaches, we conclude that the Big Four premium is still present after controlling for observable audit client characteristics and that models attributing the premium to unobservable characteristics should be treated with a high degree of caution. The Heckman procedure is becoming more widely used in accounting research such as studies of accruals quality (Doyle et al., 2007) and effects of disclosure on stock returns (Tucker, 2007). Accordingly, our results and methods are also likely to be of more general interest to accounting and business researchers. The remainder of the paper is organised as follows. In the next section, we outline general modelling issues and assumptions; Section 3 describes our variables and data, while our empirical results based on single stage, two-step and matching estimators follow in Section 4. The paper concludes in Section 5 with a summary, implications and suggestions for future research.

2. Modelling issues and the Big Four premium

2.1. Evidence on the premium in prior literature

To date, the auditing literature has advanced several non-independent reasons for large auditors charging higher fees, including the Big Four (formerly Big Eight, Big Six and Big 5) being associated with established reputations, higher quality audits, higher training costs, higher potential losses in the event of shareholder litigation ('deep pockets') and the occupation of a position of oligopoly in many audit markets (Moizer, 1997). Craswell et al. (1995) note that in competitive markets, the large auditor premium represents a return to Big Four investments in brand name reputation for higher quality audits. In the market for the largest multinational companies, however, smaller auditors, due to their lack of technical resources and geographical coverage, are unable to compete; hence such auditees are limited in choice to Big Four auditors only. For example, the Oxera report (2006: i) concludes there are significant barriers to entry in the sub-market for large UK quoted companies, 'including the high cost of entry, a long payback period for any potential investment, and significant business risks when competing against the incumbents (Big Four) in the market'.

Testing whether or not the auditee market is competitive (i.e. amongst the Big Four) for the largest companies, or subject to cartel pricing behaviour, is clearly difficult, since no realistic counterfactuals exist – with, for example, Big Four auditors accounting for 97.4% of the audits of the FTSE 350 in 2005 (Oxera, 2006). In the current paper we study UK private companies, where the market is *a priori* competitive in that Big Four concentration is relatively low (8.3% of audits in our sample) and where both Big Four and non-Big Four auditors are represented across a wide range of auditee size. In such a market any observed premium is more likely to be related to perceived or actual audit quality differentials than to a lack of competition.[3]

In our modelling of audit fees, we therefore assume a competitive market using the seminal audit fee framework of Simunic (1980) and developed by Pong and Whittington (1994). Simunic (1980) hypothesises that audit fees vary in association with audit production functions, loss exposure and audit quality (modelled with reference to auditee size, complexity, risk and auditor quality). Pong and Whittington (1994) posit that supply is related to auditors' cost functions and hence largely associated with the quantity of work/effort. Because of professional and statutory prescriptions for minimum audit standards, Pong and Whittington (1994) argue that the demand for audit is relatively inelastic. Furthermore, as noted by Simunic (1980: 170), in terms of product differentiation, the audit market is hedonic, i.e. differentiated audit products (quality) are not directly observed and 'the principal differentiation characteristic of the service is likely to be the identity of the supplier … it is the Big Eight firms which enjoy visibility and brand name recognition among buyers.'

The UK private company audit market is an interesting context in which to test for the presence of a large auditor premium. In addition to the more competitive nature of the supply side of the audit market, there are economic arguments both for and against the prediction that a Big Four premium will be observed. As argued by Chaney et al. (2004), lower agency costs for private firms (which are more closely held), potentially less reliance on financial statements by outsiders and lower litigation risk for auditors (compared to listed firms) would point to lower demand for high quality audit services, and hence to no expectation of a premium. By contrast, owners of private firms may seek to signal credibility of their financial

[3] Several studies indicate that a premium may be warranted as a result of differential audit quality. For example, Blokdijk et al. (2006) find that the quality of audits by the (then) Big Five is higher, even though the total effort exerted is similar to smaller auditors. Francis et al. (1999) report that Big Six auditors constrain income-increasing discretionary accruals more than smaller auditors, while Lennox (1999) finds that large auditors' reports are more accurate than those of smaller auditors.

Figure 1
Linear regression for audit fees

$$\ln F_{BIG4} = \alpha + \beta + \sum_{k=1}^{K} \beta_k X_{k\,BIG4} + \varepsilon_{BIG4} \qquad \text{For Big Four clients } (D = 1) \qquad (1)$$

$$\ln F_{NON} = \alpha + \beta + \sum_{k=1}^{K} \beta_k X_{k\,NON} + \varepsilon_{NON} \qquad \text{For non-Big Four clients } (D = 0) \qquad (2)$$

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

statements should they plan to sell their stake and the absence of market values may make information provided by the financial reporting process more important (e.g. for managerial performance measures). Collis et al. (2004) find increased demand for audit services in general (rather than between types of auditor) from firms wishing to maintain good relationships with lenders; it is possible that such findings may extend to firms wishing to appoint higher quality auditors to testify to the truth and fairness of their financial statements. In addition, there is evidence that newly listed firms attract cheaper debt capital if they appoint a large auditor (Pittman and Fortin, 2004), suggesting that higher audit fees may eventually be recovered through the payment of lower rates of interest.

2.2. *Statistical specifications and assumptions*

In this section, we describe the models and assumptions that provide the basis for our empirical analysis. Although our focus is on the Big Four premium, the discussion is applicable to other areas of accounting and business research where selection bias is a potential problem. At various points, we refer readers to the Appendix for a more formal exposition of the issues in this section.

OLS and Oaxaca-Blinder decomposition

We start by dividing companies into those with a Big Four auditor and those without. This division is indexed below by *BIG4* and *NON* and represented by a dummy variable (D) taking the value of one if the auditee appoints a Big Four auditor and zero otherwise. The literature typically assumes audit fees (F), expressed in natural log form ($\ln F$), depend on K variables (X_k $k = 1,..,K$) – principally auditee size, complexity and risk measures – employing a linear regression of the form shown in Figure 1, where the error term (ε) reflects unobservable random determinants of the fees paid to auditors.

Audit fees may vary between these groups because observable characteristics (X) are different and/or because the impact of these characteristics on audit fees ($\beta \neq 0$, $\alpha_k \neq \beta_k$) is different. For instance, as Pong and Whittington (1994) and Chaney et al. (2004) note, it is likely that Big Four auditors are better equipped to audit larger, more complex clients, although such comparative advantages may be offset in part by higher fixed costs associated with the training of audit staff.

Initially we assume for our single stage conventional estimates that any unobservable auditee characteristics are the same for $D = 1$ and $D = 0$, so the errors have the same distribution for each type of auditor. A problem arises since we cannot directly compare the fees paid under each regime because we only observe a company as a client of either a Big Four or a non-Big Four auditor, but not both, i.e. we do not observe the counterfactual outcome.[4] This problem can be overcome by assuming that the values of the regressors are unimportant in respect of computing the counterfactuals; however, if there are large and significant differences in the values of the regressors for Big Four and non-Big Four auditees, then it may be unreasonable to extrapolate between them.[5]

If the OLS estimates of the parameters in (1) and (2) are (a, a_k) for the non-Big Four auditees and (a, b, b_k) for the Big Four auditees, then the predicted log of audit fees for a Big Four auditee, firm i, in each audit regime are:

$$\widehat{\ln F_{NONi}} = a + \sum_{k=1}^{K} a_k X_{ki}$$

(the counterfactual value) and

$$\widehat{\ln F_{BIG4i}} = a + b + \sum_{k=1}^{K} b_k X_{ki}$$

(the actual predicted value). The Big Four premium

[4] Another potential concern is the use of linear functions. It may be possible for the same non-linear audit fee equation to apply to both types of auditee so that any observed Big Four premium might be entirely 'explained' by auditees' different characteristics. A premium can still be predicted if linear approximations are estimated at markedly different points on the curve.

[5] For example, at the limit, it would appear inappropriate to compare the audit fees paid by large and small auditees if *all* large auditees employed Big Four auditors while *all* small auditees employed non-Big Four auditors.

Figure 2
Estimates of the Big Four premium

$$P_{BIG4} = b + \sum_{k=1}^{K} (b_k - a_k)\overline{X_{kBIG4}} \quad (3)$$

$$P_{NON} = b + \sum_{k=1}^{K} (b_k - a_k)\overline{X_{kNON}} \quad (4)$$

is then the difference:

$$\widehat{\ln F_{BIG4i}} - \widehat{\ln F_{NONi}} = b + \sum_{k=1}^{K} (b_k - a_k) X_{ki}$$

(most previous studies test for a premium using a binary variable in a single regression, so the premium is constant at b). In practice we compute these statistics for two 'typical' (average) auditees: the first has the values for the regressors equal to the mean values for the Big Four auditees ($\overline{X_{kBIG4}}$) and the other the mean values for the non-Big Four auditees ($\overline{X_{kNON}}$).[6] This gives two estimates (P) of the Big Four premium, shown in Figure 2.

P_{BIG4} represents the predicted fees paid by a typical Big Four auditee to a Big Four auditor minus the predicted log of fees paid by the same auditee to a non-Big Four auditor; whereas P_{NON} represents the difference between predicted fees for a typical non-Big Four auditee paid to a Big Four auditor and the predicted fees paid to a non-Big Four auditor. Although not typically used in auditing research, these statistics are widely used elsewhere as part of an Oaxaca-Blinder (OB) decomposition analysis (see the Appendix, Oaxaca (1973), Blinder (1973) and Greene (2003: 53) for further details).

Recent developments in the auditing literature, however, point out that conventional OLS estimates of the Big Four premium are potentially biased since auditors are not appointed randomly by their clients and because auditor choice may be systematically related to auditees' unobservable characteristics (e.g. insiders' knowledge of the riskiness of future cash flows). Ireland and Lennox (2002: 75) note 'although the standard OLS audit fee models control for observable differences, characteristics that are not observable to the academic researcher may affect both fees and auditor choice and thereby cause bias.' However, it is not entirely clear from previous research what such unobservable characteristics represent or how important they are in systematically influencing auditor selection and audit fees. Titman and Trueman (1986) and Datar et al. (1991) each develop models predicting that auditor quality is a function of firm-specific risk, of which firm insiders are better informed than outsiders. But both models make competing predictions about the nature of the relationship between firm-specific risk and auditor quality: Datar et al. (1991) predict that entrepreneurs of risky firms choose higher quality auditors, whereas Titman and Trueman (1986) predict the opposite.

Selection models

Selection bias arises if the unobservable characteristics of Big Four and non-Big Four auditees are systematically different from each other. Suppose that ε_{NON} and ε_{BIG4} in equations (1) and (2) above are drawn from the same distribution but that Big Four auditees and non-Big Four auditees only have positive and negative errors respectively.[7] Estimating fee equations with standard single-stage OLS omits the conditional means (by assuming $E(\varepsilon_{BIG4}) = E(\varepsilon_{NON}) = 0$) and leads to inconsistent estimates if these terms are correlated with the regressors. The Heckman two-step procedure provides an estimate of the mean of the conditional error known as the Inverse Mills Ratio (IMR) or the selection term λ, which augments the regressors in equations (1) and (2) above. The procedure involves estimating a probit model of auditor choice as the first stage; this model yields estimates of selection terms λ_{BIG4} and λ_{NON} which are then included in the audit fee equations in the second stage. OLS applied to the augmented equations yields consistent coefficient estimates and standard hypothesis tests can be applied with modified formulae for the standard errors. The Heckman procedure thus estimates the equations shown in Figure 3.

The Appendix provides a more formal description of the model and the derivation of the selection terms. The probit and audit fee equations in

[6] This choice ensures that the errors play no role as the means of the predicted errors are zero and would seem reasonable on the basis that the mean represents the expected value of the characteristics of a Big Four auditee.

[7] For example, if the positive error measures the unobserved value to the auditee of appointing a Big Four auditor, then Big Four auditees will value Big Four auditors more than non-Big Four auditors and therefore pay higher fees.

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Figure 3
Equations for Heckman procedure

$$\ln F_{BIG4} = \alpha + \beta + \sum_{k=1}^{K} \beta_k X_{k\,BIG4} + \sigma_{SBIG4}\lambda_{BIG4} + \upsilon \qquad \text{For Big Four auditees} \qquad (5)$$

$$\ln F_{NON} = \alpha + \sum_{k=1}^{K} \alpha_k X_{k\,NON} - \sigma_{SNON}\lambda_{NON} + \upsilon \qquad \text{For non-Big Four auditees} \qquad (6)$$

the Heckman model can also be estimated simultaneously by maximum likelihood (ML), which leads to more efficient estimates if the model is correctly specified. Accordingly, we report both conventional Heckman two-step and ML estimates in our empirical analysis.

Although the Heckman procedure has become increasingly popular in accounting research and in the finance literature (e.g. Li and Prabhala, 2007), its robustness has been questioned under certain conditions. For example, Giles (2003: 1299) notes 'Heckman's sample selectivity correction methodology offers a way of improving on the estimates obtained with non-random samples. While there is improvement in general in this regard, there are situations in which the correction for sample selectivity actually aggravates the problem.'

It is commonplace to assume joint normality of the distribution of the errors in the selection and outcome equations and that any systematic unobservable variables are normally distributed (an untestable assumption). Joint normality has the surprising and unfortunate implication that collinearity between the selection term and the other regressors in the second stage equation is often severe, leading to serious model instability (e.g. Leung and Yu, 2000). In addition, it is common for researchers to identify the second stage equation via the non-linearity of the selection term only. However, recent econometric analyses of this issue suggest that to adequately identify the model it should contain a valid instrument, i.e. a regressor which determines the choice of auditor but has no significant effect on determining audit fees (Little, 1985; and Puhani, 2000). But collinearity may cause problems even when an instrument (also known as an exclusion or identifying variable) is employed, leading to unstable estimates of treatment effects (Stolzenberg and Relles, 1997; Leung and Yu, 2000; Li and Prabhala, 2007). Against this background, it is perhaps unsurprising that, as noted above, empirical results in auditing research using the Heckman model have, to date, been mixed.[8]

In the absence of satisfactory instruments, the selection effect is best identified by extreme observations of the selection term λ, i.e. those companies with an estimated probability of choosing a Big Four auditor close to 1. These Big Four auditees (usually because of their large size and complexity) effectively have no surrogate non-Big Four counterfactuals – that is, there is no 'common support' – the common support region being where Big Four clients have non-Big Four counterparts with similar characteristics. Following Black and Smith (2004), we therefore assess the robustness of the Heckman results by estimating models using samples with different values of the selection term.

Matching estimators

The problems of model sensitivity, lack of robustness, linear functional form assumptions and the need for adequate counterfactuals motivate the use of matching methods. These methods are gaining in popularity in the applied econometrics literature (e.g. Bryson et al., 2002; Black and Smith, 2004; Simonsen and Skipper, 2006) and are based on matching the observable characteristics of members in the treatment group (i.e. Big Four auditees in our case) to members (counterfactuals) in the untreated group (non-Big Four auditees). Matching analyses are based on two important assumptions: (1) the conditional independence assumption (CIA) and (2) common support. The former requires the value of audit fees to be independent of auditor type given the values of some observable variables, whereas the latter involves comparable observations existing in both groups. Both assumptions are discussed at greater length in the Appendix.

A limitation of matching methods is that they cannot accommodate any systematic effects of unobservable auditee characteristics that have a joint impact upon auditor selection and audit fees (although as discussed above, it is not obvious what these systematic effects might be or what their directional influence is). Whereas the Heckman

[8] Though it is, of course, plausible that the variation in these findings is attributable to differences in the underlying economic situations (e.g. private limited versus public quoted company markets).

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

approach allows such unobservable factors to influence auditor choice systematically, it is often sensitive to specification and collinearity, as discussed above. Although they cannot deal directly with unobservables, matching estimators do not rely on linear extrapolation (outside the common support region) or functional form assumptions; nor do they require an exclusion variable or impose joint normality assumptions. As noted by Simonsen and Skipper (2006), matching methods are based on matching on observable characteristics of members in the treatment group (Big Four auditees in our research) to members (counterfactuals) in the untreated group (non-Big Four auditees) and hence 'balancing the bias arising from self-selection … Matching allows for heterogeneous treatment effects, is not subject to parametric assumptions and does not *per se* assume separability of observables and unobservables' (*ibid*.: 920).

When applying matching methods, there are various estimators to choose from, reflecting a trade-off in respect of the number of variables used to match on, the closeness with which the variables are matched (particularly continuous variables), and the sample size. This is referred to as the 'curse of dimensionality' (Ho et al., 2007) since close matching on more than a few variables (dimensions) might result in matched samples that are too small for any meaningful analysis.

The first matching method we use – propensity score matching – has recently been employed in applied econometrics research (e.g. Black and Smith, 2004; Simonsen and Skipper, 2006), but not in the auditing literature to date. The method (see Appendix for more details) is implemented as follows. First the selection equation is estimated using a parametric estimator of the auditor selection equation (in our case a probit model) and the probabilities (propensity scores) of choosing a Big Four auditor are obtained for all sample firms. Each Big Four auditee is then matched to a non-Big Four auditee with a similar propensity score and differences in audit fees compared across the two matched samples. As noted by Black and Smith (2004: 110), the logic underpinning this method is that 'subgroups with values of X [explanatory variables] that imply the same probability of treatment can be combined because they will always appear in the treatment and (matched) comparison groups in the same proportion. As a result, any differences between subgroups with different X but the same propensity score balance out when constructing the estimates.' Hence, an important practical advantage of propensity score matching is that subgroups are matched on one variable, obviating the need for very large samples when subgroups are to be matched according to several characteristics. Moreover, propensity score matching does not make the same functional form assumptions as linear regression and non-linear relationships (of which there is some evidence in the literature – e.g. Peel and Roberts, 2003; Chaney et al., 2005) can be allowed for.

The second matching approach we take is an intermediate (semi-parametric) one which combines matching with the standard OLS regression used in the majority of prior studies. This approach involves matching observations according to important *actual* client characteristics prior to a standard parametric analysis, in line with the methods advocated by Ho et al. (2007), who highlight the potentially serious pitfalls of drawing inferences from sensitive statistical models. Initially we preprocess our data, then estimate the standard audit-fee model with a binary Big Four indicator variable. Preprocessing involves matching Big Four and non-Big Four auditees only on key attributes (well-tested measures of auditee size, complexity and risk) thereby ensuring a sufficient number of matched observations to conduct standard OLS regression techniques to control for any remaining confounding factors. We initially sort our full sample into quantiles based on the key attributes of sales (40 quantiles), the ratio of exports to sales (11 quantiles), return on total assets (40 quantiles) and the number of subsidiaries (10 quantiles). We then match each Big Four client to a non-Big Four client jointly sharing membership of each respective quantile for size, risk and complexity. It is important to note that this is a simultaneous requirement, i.e. matched auditees have similar size *and* risk *and* complexity characteristics and is therefore a stricter set of criteria than propensity score matching, since the latter is a composite (albeit conditional) score.

Since this process results in a large number of potential matched combinations, we perform this procedure 2,000 times. Each time, we estimate an OLS regression and for each iteration, we capture the coefficient for the binary Big Four indicator variable (representing the premium charged to similar Big Four and non-Big Four auditees) and report the results for the distribution of this coefficient. As stated by Ho et al. (2007: 3) this pre-processing approach combines the merits of both non-parametric matching with conventional parametric estimators: 'In a sense our recommendations already constitute best practice since matching alone is not a method of estimation and always requires some technique to compute estimates … we simply point out that, except in the extraordinary case where matching is exact, parametric procedures have the potential to greatly improve causal inferences even after matching.' In summary, while the Heckman model presents a potential solution to the important problem of selection bias, it may produce imprecise results.

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Table 1
Variable definitions

Label	*Definition*
ln*AFEE*	Natural log of audit fee (in £)
ln*SAL*	Natural log of turnover (in £)
ln*TA*	Natural log of total assets (in £)
SQSUBS	Square root of the number of subsidiaries
EXPSAL	Ratio of non-UK turnover to total turnover
QUALIF	Binary variable taking the value of 1 if company had qualified audit report, 0 otherwise
PBAL	Binary variable taking the value of 1 if company disclosed a post-balance sheet event in accounts, 0 otherwise
CONLIAB	Binary variable taking the value of 1 if company disclosed contingent liabilities in accounts, 0 otherwise
EXITEM	Binary variable taking the value of 1 if company disclosed exceptional and/or extraordinary items in accounts, 0 otherwise
RTA	Ratio of profit before tax to total assets
TLTA	Ratio of total liabilities to total assets
LOND	Binary variable taking the value of 1 if company is located in London, 0 otherwise
BUSY	Binary variable taking the value of 1 if firm's year-end is in December or March, 0 otherwise
BIG4	Binary variable taking the value of 1 if company is audited by a Big Four auditor, 0 otherwise
CHTA	Absolute value of change in total assets from year *t*-1 to year *t*

Matching techniques are becoming increasingly popular, since as noted by Li and Prabhala, (2007: 51), 'They represent an attractive means of inference because they are simple to implement and yield readily interpretable estimates of "treatment effects "'. However, they are based on different assumptions to the Heckman model, principally because they assume that any unobservables are unimportant (Li and Prabhala, 2007). If unobservable client characteristics determine both auditor choice *and* audit fees, matching estimators produce potentially biased results. Since the theoretical research into the determinants of auditor choice is inconclusive and the nature and role of unobservables are unclear, however, matching methods seem an appropriate means to assess the robustness of the Big Four premium in recent auditing research.[9] The next section outlines the variables and data used in our empirical analysis.

3. Variables and data

3.1. Variables

Our main empirical model of audit fees takes the standard linear form as in Figure 4.

The variables used in the model are described in Table 1 and have been widely employed in prior research (e.g. Simunic, 1980; Pong and Whittington, 1994; Chan et al., 1993; Ezzamel et al. 1996; Chaney et al., 2004; McMeeking et al., 2006; Clatworthy and Peel, 2007).[10]

Since corporate size (serving as a proxy for audit effort) has been found to be the key driver of external audit fees in previous research, we employ both total assets (£) and turnover (£) as auditee size measures in our research. Pong and Whittington (1994: 1075) note that audits have two broad dimensions: 'an audit of transactions and verification of assets. The former will be related to turnover and the latter to total assets.'

[9] Since prior research on this specific issue is relatively rare and has produced inconsistent results, further research into the identification and examination of such unobservable characteristics seems warranted.

[10] Naturally, the binary variable *BIG4* is omitted where the equations are estimated separately. As noted by an anonymous referee, a potentially important variable not included in our analysis is non-audit (consultancy) fees. This variable has been found to be significantly related to audit fees in a number of studies of listed companies; however, since private companies are not required to disclose their non-audit fees, the data are not available for most private companies (i.e. other than for those that voluntarily disclose them). This is a potential limitation of our research, since the distinction between the two types of fees is not always clear – though previous Heckman (two-step) research into UK private firms also omits non-audit fees, so our results are comparable in this respect.

Figure 4
Empirical model of audit fees

$$\ln AFEE = \alpha_0 + \beta_1 \ln SAL + \beta_2 \ln TA + \beta_3 SQSUBS + \beta_4 EXPSAL + \beta_5 QUALIF + \beta_6 PBAL + \beta_7 CONLIAB + \beta_8 EXITEM + \beta_9 RTA + \beta_{10} TLTA + \beta_{11} LOND + \beta_{12} BUSY + \beta_{13} BIG4 + \varepsilon \quad (7)$$

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Following previous studies, we specify the relationship between audit fees (ln*AFEE*) and the size measures for turnover (ln*SAL*) and total assets (ln*TA*) in natural logarithmic form to capture potential economies of scale in the audit. In order to control for audit complexity, we include a variable labelled *SQSUBS*, defined as the square root of the number of subsidiaries (e.g. Francis and Simon, 1987), and *EXPSAL* – the ratio of non-UK turnover to total turnover (e.g. Beatty, 1993: Chaney et al., 2004), both of which we expect to be positively related to audit fees.

To capture auditee risk characteristics, we employ the gearing ratio of total liabilities to total assets (*TLTA*) and the ratio of net profit before tax to total assets (*RTA*), which we expect to be positively and negatively related to audit fees, respectively (e.g. Chan et al., 1993; Firth, 1997). Following previous research (e.g. Chaney et al., 2004; Clatworthy and Peel, 2007) we employ three additional binary variables to capture incremental risk/complexity in the audit. These are whether (coded 1) or not (coded 0) the audit client received a qualified audit report (*QUALIF*), reported exceptional and/or extraordinary items (*EXITEM*), disclosed a post-balance sheet event (*PBAL*) or a contingent liability (*CONLIAB*). All these variables are expected to be positively related to audit fees (*ibid*.).[11] Finally, we include binary variables for whether (coded 1) or not (coded 0) companies are audited by a Big Four auditor (*BIG4*), whether the audit client's year-end falls in December or March (*BUSY*) and whether the company is located in London (*LOND*). The latter two variables are expected to be positively related to audit fees since companies audited during the 'busy' period may be charged higher fees due to the higher opportunity cost of audit resources (e.g. McMeeking et al., 2006) while companies located in London are expected to pay higher audit fees reflecting cost of living differentials (Chaney et al., 2004; Clatworthy and Peel, 2007).

Other than in respect of corporate size and complexity, the literature on the choice of variables in the auditor selection model is less developed and prior studies are usually based on including a subgroup of variables from the audit fee equation in the selection model (Chaney et al., 2004, 2005; Hamilton et al., 2005, but cf. Ireland and Lennox, 2002). If one assumes that firms choose auditor type by comparing their predicted costs (fees), the choice of auditor type depends on all the factors affecting the fees charged by either type of auditor. We therefore included all variables from the fee equations in the auditor choice model. When all or a subset of the regressors from the fees equation is used, identification relies on the non-linearity of the selection term but this non-linearity may not be sufficient to produce convincing estimates. It is important therefore to include an identification variable that is significantly associated with auditor choice (in the probit model), but not with audit fees (in the fees equation). Such variables are extremely hard to obtain in practice (see, e.g. Puhani, 2000). We considered several plausible instruments[12] and found only one – the change in the absolute value of total assets (*CHTA*) between the current and preceding year – which was statistically significant (with the expected sign) in the probit selection model, but statistically insignificant when included in the OLS audit fee models.[13] Furthermore it is not formally a 'weak' instrument since it has an *F*-statistic of 11.21 for the null that it is insignificant in the regression of auditor type (*D*) on all the regressors. This exceeds the critical value of 8.96 for the validity of a single instrument given by Stock et al. (2002) and the informal value of 10 that is widely used and advocated by Stock and Watson (2003: 350).

The motivation for *CHTA* being included in the selection model is that companies involved in large investments/acquisition or divestments/sale of assets, may require the expertise of a Big Four auditor due to the additional complexity of the audit. In addition, Keasey and Watson (1991) note that the absolute change in firm size (total assets) may, from an agency perspective, act as a proxy for contractual changes at the firm level, which

[11] Because company records on the database we use (FAME) only indicate whether or not either of these events occurred, we are unable to refine *PBAL* or *CONLIAB* to take account of the types of events or the nature of liabilities. Hence, we assess their average impact. We were also unable to ascertain the nature of the qualification and hence through *QUALIF*, we again measure the average impact of a qualified audit report.

[12] These included changes in sales, change in equity, change in total assets and various transformations of these variables.

[13] The *t*-values for the *CHTA* coefficients when included in models 1, 2 and 3 in Table 3 below were, respectively: 0.05, 1.32 and 0.18.

Figure 5
Empirical model of auditor selection

$$BIG4 = \delta_0 + \delta_1 \ln SAL + \delta_2 \ln TA + \delta_3 SQSUBS + \delta_4 EXPSAL + \delta_5 QUALIF + \delta_6 PBAL + \delta_7 CONLIAB + \delta_8 EXITEM + \delta_9 RTA + \delta_{10} TLTA + \delta_{11} LOND + \delta_{12} BUSY + \delta_{13} CHTA + \varepsilon \quad (8)$$

could prompt a change in the demand for auditing services. Hence, large auditors may be associated with reducing agency costs (e.g. Ireland and Lennox, 2002) in companies with large asset variations. Although it has desirable theoretical qualities, it is also employed for pragmatic reasons, since it formally fulfils its main purpose of properly identifying the audit fee equations. Our empirical model of auditor selection is shown in Figure 5.

Following previous studies (e.g. Chaney et al., 2004; Hamilton et al., 2005), we expect the variables reflecting auditee size (ln*SAL* and ln*TA*) and complexity (*SQSUBS* and *EXPSAL*) to be positively associated with the choice of a Big Four auditor in the probit model, in consequence of their hypothesised capacity to provide more efficient audits and to reduce agency costs (ibid.). In line with prior research (e.g. Ireland and Lennox, 2002; Chaney et al., 2004, 2005), we also expect our auditee risk variables (*QUALIF*, *PBAL*, *CONLIAB*, *EXITEM*, *RTA* and *TLTA*) to be positively associated with the selection of a Big Four auditor.[14] As noted by Hamilton et al. (2005: 9), 'The greater the client's risk, the higher the propensity for the impairment of agency relationships. To mitigate the associated agency costs, higher quality auditors, surrogated by big 4, are more likely to be selected to signal the credibility of reporting.' Furthermore, Datar et al. (1991) predict, and Copley and Douthett (2002) find, a positive relationship between auditee risk and the appointment of a higher quality auditor.

For the final two variables (*LOND* and *BUSY*), we have no strong prior expectations about their influence on auditor choice, although the univariate results of extant studies (as in the current study) have consistently reported (for both private and quoted audit clients) that a significantly higher proportion of Big Four auditors conduct their audits during the busy period, while a significantly higher proportion of non-Big Four auditors are appointed to companies located in London (e.g. Ireland and Lennox, 2002; Chaney et al., 2004).

3.2. Data

The source of our data is the Bureau Van Dijk FAME DVD-ROM UK database. Financial data (annual accounts) and non-financial data (e.g. company location, auditor and audit qualification) are available as individual records for each company on the database. Companies were included if they met the following criteria: their primary activities (according to FAME primary SIC codes) were outside the financial sector; they were private limited companies; they were 'live' companies (i.e. had not ceased trading, failed or entered into voluntary liquidation); their audited accounts were available on FAME; they had full data available, including total assets and sales (minimum £1,000), audit fee (minimum £100), and a disclosed profit/loss figure. In order to avoid the potential confounding influences of including both holding companies and their subsidiaries in the regression model (e.g. Ezzamel et al., 1996; Peel and Roberts, 2003), our sample only includes independent companies (i.e. those not held as a subsidiary of another company). In line with previous studies (e.g. Firth, 1997), financial companies were excluded due to the different composition of their financial statements and only live companies were selected to avoid the confounding influence of including non-live auditees. In addition, and in line with previous research, 11 companies with joint auditors (none of which were Big Four auditors) were excluded from the analysis to comply with the binary nature of the probit model. Following these restrictions, we obtained the necessary data for a sample of 36,674 private companies from FAME for the latest financial statements available (predominantly for the calendar year 2003).[15] It is very important to note (since it has significant effects on both sample size and data accuracy) that the FAME default setting for downloading data is £000s, with data being rounded to the nearest £1,000; for example an audit fee of £1,550 would be rounded to £2,000 and one of £400 to zero (i.e.

[14] As noted by an anonymous reviewer, it is possible that if large auditors do not favour small companies due to lower return relative to the risk involved, small firms will have to pay more to retain a large auditor. This does not, however, preclude small companies paying such a premium if they believe the quality of audit to be higher for the reasons discussed above.

[15] A possible limitation of our work is that we employ a cross-sectional, rather than a panel-based sample. Prior research by Chaney et al. (2004) suggests that private company audit fee models do not vary much over time. Their sample covers a five-year period and there is little variation in their annual estimates; however, cf. McMeeking et al. (2006), who do find inter-temporal variation for listed companies.

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Table 2
Descriptive statistics

	Big Four clients (n = 3,038)			*Non-Big Four clients (n = 33,636)*			*Total sample (n = 36,674)*			
Variable	*Mean*	*Std. dev.*	*Median*	*Mean*	*Std. dev.*	*Median*	*Mean*	*Std. dev.*	*Median*	*Sig.*
AFEE (£000)	29.05	80.47	13.00	5.88	13.12	2.75	7.80	27.11	3.00	‡§
ln*AFEE*	9.44	1.25	9.47	7.91	1.22	7.92	8.04	1.29	8.01	‡§
SAL (£m)	39.41	150.62	8.14	5.13	21.46	0.84	7.97	48.89	1.02	‡§
TA (£m)	35.62	159.41	6.08	3.17	14.49	0.48	5.86	48.75	0.59	‡§
SUBS	3.46	8.37	1.00	0.61	2.46	0.00	0.84	3.46	0.00	‡§
EXPSAL	0.08	0.20	0.00	0.02	0.12	0.00	0.03	0.13	0.00	‡§
QUALIF	0.04	0.19	0.00	0.03	0.17	0.00	0.03	0.18	0.00	
PBAL	0.12	0.33	0.00	0.04	0.19	0.00	0.04	0.21	0.00	ψ
CONLIAB	0.27	0.45	0.00	0.10	0.30	0.00	0.11	0.31	0.00	ψ
EXITEM	0.57	0.50	1.00	0.32	0.47	0.00	0.34	0.47	0.00	ψ
RTA	0.01	0.38	0.03	0.23	0.78	0.07	0.21	0.76	0.07	‡§
TLTA	0.84	1.51	0.73	0.77	1.15	0.66	0.78	1.18	0.67	‡§
LOND	0.21	0.41	0.00	0.34	0.47	0.00	0.33	0.47	0.00	ψ
BUSY	0.56	0.50	1.00	0.44	0.50	0.00	0.45	0.50	0.00	ψ
CHTA (£m)	5.26	33.82	0.61	0.49	2.71	0.05	0.88	10.16	0.06	‡§

Notes:
Variable definitions are provided in Table 1.
‡ and § indicate means and distributions are significantly different between Big Four and non-Big Four clients at the 0.01 level in *t*-tests and Mann-Whitney tests respectively.
ψ indicates significant difference between Big Four and non-Big Four clients at the 0.01 level in a chi-squared test.

a missing value). Data can, however, be downloaded (as in the current study) in £ and hence neither data accuracy nor observations are lost using this option. The sampling consequences of this are not trivial, since downloading in £ captures a large number of smaller firms. For instance, Chaney et al. (2004), whose sample excludes many small companies due to the imprecision associated with downloading in £000, report Big Four concentration of 50% compared to 8% in our sample.

Descriptive statistics are presented in Table 2. The average audit fee (*AFEE*) for the whole sample (n = 36,674) amounted to £7.80k, with companies having mean sales (*SAL*) and total assets (*TA*) of £7.97m and £5.86m respectively. Sales range from a minimum of £1k to a maximum of £4,979m and total assets from £1k to £5,234m. Table 2 also shows that, other than audit qualifications (*QUALIF*), all variables differ significantly between the Big Four (n = 3,038) and the non-Big Four (n = 33,636) sub-samples. Consistent with prior expectations, Big Four clients are significantly larger (as measured by both *SAL* and *TA*), have more subsidiaries (*SUBS*), have a higher proportion of foreign to total sales (*EXPSAL*) and report more post balance sheet events (*PBAL*), contingent liabilities (*CONLIAB*) and exceptional items (*EXITEM*). In addition, Big Four clients are less profitable (*RTA*) more highly geared (*TLTA*), less likely to be located in London (*LOND*), more likely to be audited during the busy period (*BUSY*), with a significantly higher absolute change in the value of total assets (*CHTA*). Due to the large number of small auditees represented in the non-Big Four sample, the differences in size between Big Four (average sales and total assets of £39.41m and £35.62m) and non-Big Four auditees (average sales and total assets of £5.13m and £3.17m) are substantial.

4. Empirical results

We commence our analysis with standard single-stage OLS regression under the assumption of no selection bias. We then report our comparative analysis employing the two-step Heckman procedure, together with associated robustness tests. Finally, we present the results of the matching procedures.

4.1. Single stage results

Model 1 in Table 3 shows the OLS estimates for the standard pooled audit fee specification, which is employed in many previous studies. All explanatory variables take their expected signs and other than the busy period variable (*BUSY*), which is statistically significant at the 0.10 level (p = 0.079), all are highly significant (p < 0.0001 in all cases).

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Table 3
Regression results

	OLS single stage models			MLE two-step models			Heckman two-step models		
	Model 1 (Pooled fee)	*Model 2 (Big Four fee)*	*Model 3 (Non-Big Four fee)*	*Model 4a (Probit choice)*	*Model 5a (Big Four fee)*	*Model 6a (Non-Big Four fee)*	*Model 4b (Probit choice)*	*Model 5b (Big Four fee)*	*Model 6b (Non-Big Four fee)*
ln*SAL*	0.284	0.285	0.286	–0.004	0.285	0.286	–0.005	0.285	0.285
	(87.31)**	(25.08)**	(83.84)**	(0.48)	(30.93)**	(96.44)**	(0.50)	(29.51)**	(92.71)**
ln*TA*	0.122	0.120	0.120	0.246	0.149	0.113	0.246	0.211	0.101
	(36.28)**	(10.30)**	(33.54)**	(23.25)**	(8.00)**	(34.86)**	(23.20)**	(5.41)**	(25.46)**
SQSUBS	0.258	0.201	0.281	0.087	0.209	0.268	0.088	0.227	0.247
	(44.70)**	(19.07)**	(40.10)**	(6.98)**	(19.81)**	(46.05)**	(7.03)**	(15.28)**	(34.78)**
EXPSAL	0.367	0.627	0.293	0.607	0.688	0.253	0.602	0.822	0.191
	(13.15)**	(10.91)**	(9.41)**	(9.84)**	(11.34)**	(8.98)**	(9.77)**	(8.36)**	(6.17)**
QUALIF	0.115	0.141	0.111	–0.146	0.126	0.118	–0.146	0.094	0.128
	(6.18)**	(2.56)*	(5.58)**	(2.61)**	(2.30)*	(6.06)**	(2.61)**	(1.57)	(6.36)**
PBAL	0.119	0.179	0.098	0.169	0.196	0.084	0.169	0.233	0.063
	(7.74)**	(5.65)**	(5.62)**	(4.23)**	(5.85)**	(4.69)**	(4.23)**	(5.62)**	(3.34)**
CONLIAB	0.095	0.064	0.099	0.014	0.064	0.096	0.013	0.066	0.091
	(8.93)**	(2.64)**	(8.47)**	(0.48)	(2.57)*	(7.68)**	(0.45)	(2.45)*	(7.09)**
EXITEM	0.131	0.126	0.130	–0.081	0.118	0.133	–0.080	0.100	0.137
	(17.04)**	(5.57)**	(15.88)**	(3.38)**	(5.06)**	(16.14)**	(3.35)**	(3.78)**	(15.99)**
TLTA	0.026	–0.009	0.029	0.076	–0.004	0.025	0.076	0.007	0.019
	(6.43)**	(1.31)	(6.24)**	(9.07)**	(0.48)	(8.04)**	(8.98)**	(0.59)	(5.38)**
RTA	–0.033	–0.111	–0.031	–0.231	–0.139	–0.031	–0.230	–0.199	–0.032
	(7.14)**	(3.28)**	(6.65)**	(8.12)**	(3.87)**	(6.82)**	(8.10)**	(4.02)**	(6.54)**
LOND	0.208	0.338	0.200	–0.306	0.306	0.210	–0.305	0.237	0.226
	(29.69)**	(12.38)**	(27.63)**	(12.15)**	(9.93)**	(29.10)**	(12.11)**	(4.72)**	(27.93)**
BUSY	0.011	0.010	0.010	0.153	0.026	0.004	0.152	0.0601	–0.005
	(1.76)	(0.52)	(1.51)	(7.12)**	(1.18)	(0.63)	(7.10)**	(2.01)*	(0.65)

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Table 3
Regression results *(continued)*

	OLS single stage models			MLE two-step models			Heckman two-step models		
	Model 1 (Pooled fee)	*Model 2 (Big Four fee)*	*Model 3 (Non-Big Four fee)*	*Model 4a (Probit choice)*	*Model 5a (Big Four fee)*	*Model 6a (Non-Big Four fee)*	*Model 4b (Probit choice)*	*Model 5b (Big Four fee)*	*Model 6b (Non-Big Four fee)*
CONSTANT	2.299 (88.49)**	2.638 (23.07)**	2.302 (83.80)**	–4.886 (46.58)**	1.967 (5.14)**	2.378 (87.60)**	–4.870 (46.62)**	0.511 (0.58)	2.499 (69.11)**
BIG4	0.270 (22.96)**								
CHTA				0.007 (4.17)**			0.007 (4.27)**		
IMR (λ)					0.142 (1.80)	–0.199 (9.56)**		0.446 (2.42)**	–0.509 (8.39)**
N	36,674	3,038	33,636	36,674	3,038	33,636	36,674	3,038	33,636
Adj. R^2	0.78	0.80	0.75	–	–	–	–	0.80	0.75

Notes: This table reports regression estimates where the dependent variable is lnAFEE for all models, except Models 4a and 4b where the binary dependent variable is coded 1 if the firm is audited by a Big Four firm, 0 otherwise. Variable definitions are provided in Table 1. Absolute *t*-statistics are reported in parentheses, except for Models 4b–6b, where *z*-statistics are in parentheses; *t*-statistics for OLS models use robust standard errors and corrected standard errors for the Heckman two-step models. * and ** indicate statistical significance at 0.05 and 0.01 levels respectively.

Figure 6
OB decomposition

$$\overline{\ln F_{BIG4}} - \overline{\ln F_{NON}} = \sum_{k=1}^{K} a_k(\overline{X_{BIG4k}} - \overline{X_{NONk}}) + b + \sum_{k=1}^{K}(b_k - a_k)\overline{X_{BIG4k}} \quad (9)$$

Actual difference = Explained by characteristics + Big Four premium

9.4364 – 7.9099	=	9.1809 – 7.9099	+	9.4364 – 9.1809	(10)
1.5265	=	1.2709	+	0.2556	
		(263.6)		(21.87)	

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

In particular, we note that the *BIG4* coefficient (0.270) implies that, on average, the audit fees of a non-Big Four auditee would increase by 31% if it were to employ a Big Four auditor.[16] Also noteworthy is that the model explains a relatively high proportion (R^2 of 78%) of the variation in the audit fees of UK private companies, comparing favourably with that (57%) reported by Chaney et al. (2004) for their sample of UK private firms.

Models 2 and 3 in Table 3 report OLS estimates for separate audit fee equations for the Big Four and non-Big Four auditee samples. In Model 1, the Big Four equation only differs by a constant from the non-Big Four equation. In common with Chaney et al. (2004) a joint F-test rejected the null hypothesis ($F = 13.43$; $p = 0.000$) that the coefficients in the Models 2 and 3 were the same, implying that the fee-setting process differs between the two auditor types. The main focus of our empirical analysis is therefore Models 2 and 3 in Table 3 (i.e. those which allow the slope coefficients of the explanatory variables to differ for the Big Four and non-Big Four models).

Table 3 shows that for the non-Big Four specification (Model 3) all explanatory variables exhibit their expected signs and, other than for *BUSY*, which loses statistical significance ($p = 0.131$), all variables are highly significant ($p < 0.0001$ in all cases). For the Big Four specification (Model 3), in addition to *BUSY*, the sign on the gearing coefficient (*TLTA*) is negative, but statistically insignificant – a finding in common with Chaney et al. (2004) for their Big Four equation; furthermore, the intercept in Model 2 is larger than in Model 3 – a result also reported by Chaney et al. (2004) and attributed to Big Four auditors recovering higher expenditure on training and facilities.

To examine the premium in more detail, we conduct the Oaxaca-Blinder decomposition (discussed above) on our estimates for Models 2 and 3. The OB decomposition is based on measuring the premium (using the characteristics of the average Big Four auditee) as in Figure 6.

Greene (2003: 54) provides the formulae for the estimated standard errors of each term in the decomposition and we report the *t*-values based on this method in parentheses under the estimates. There is a large and significant ($p = 0.000$) difference in the means of the audit fees paid by companies audited by Big Four and non-Big Four auditors (1.5265) using the parameters from Models 2 and 3. Most of this is accounted for by differences in their respective client characteristics (1.2709 or 83%). However, there is, on average, a significant ($p = 0.000$) Big Four premium of 0.2556 (29.1%), which is close to that (31.0%) estimated in the pooled OLS equation (Model 1) and in line with findings in prior research (e.g. Moizer, 1997). On average, Big Four auditees paid audit fees of £12,537 ($e^{9.4364}$), but would have paid £9,710 if they were charged according to the non-Big Four parameters (Model 3) – a reduction of 23%.[17] Hence, the Oaxaca-Blinder results, based on Models 2 and 3 are consistent with the presence of a Big Four audit premium. The next section presents our two-stage results where we analyse the extent to which these findings are affected by selection bias.

4.2. Heckman two-step regressions

Table 3 reports the two-step results with maximum likelihood estimates (MLE) and standard Heckman two-step estimates. Models 4a and 4b show the probit selection model estimates for the choice of a Big Four auditor, while Models 5a (5b) and 6a (6b) report the MLE (standard Heckman) audit fee regression estimates for the Big Four and non-Big Four auditees, including the additional parameter λ (for the IMR estimated from the coefficients in Model 4) to control for selection bias. The MLE and standard Heckman two-step probit selection models (4a and 4b) are very similar, and in both models, all explanatory variables other

[16] We use the standard transformation $e^x - 1$ (where x = the coefficient or mean log difference) to compute percentages.

[17] The alternative decomposition using the characteristics of non-Big Four clients also implied a statistically significant premium (at $p = 0.000$) of 31%.

Figure 7
Predicted fees

$$\overline{\ln F_{BIG4}} = a + b + \sum_{k=1}^{K} b_k \overline{X_{k\,BIG4}} + g_{SBIG}\overline{\lambda_{BIG4,\,BIG4}} \qquad \text{For Big Four clients (actual)} \qquad (11)$$

$$\overline{\ln F_{NON}} = a + \sum_{k=1}^{K} a_k \overline{X_{k\,BIG4}} + g_{SNON}\overline{\lambda_{BIG4,\,BIG4}} \qquad \text{For Big Four clients (counterfactual)} \qquad (12)$$

than *CONLIAB* and ln*SAL* are significantly associated with auditor choice at the 0.01 level. In particular, the coefficient on the identifying variable (*CHTA*) exhibits its expected sign and is highly statistically significant ($p = 0.000$).[18] Also consistent with expectations and prior research, Models 4a and 4b show that larger (ln*TA*), more complex (*SQSUBS*; *EXPSAL*) and riskier (*RTA*; *TLTA*) companies are more likely to appoint a Big Four auditor. Companies receiving audit qualifications (*QUALIF*) are more likely to employ a non-Big Four auditor, in contrast to companies reporting a post-balance sheet event (*PBAL*) and auditees based in London, which are less likely to select a Big Four auditor, although likely to do so if their year ends fall in the busy period (*BUSY*).[19]

The audit fee equations (Models 5a, 6a, 5b and 6b) contain the same pattern of significance levels as the single stage estimates in Models 2 and 3. The MLE estimates in Table 3 show that the λ coefficient is negative but highly significant ($p = 0.000$) for the non-Big Four equation (Model 6a), but positive and only significant at the 0.10 level ($p = 0.071$) in the Big Four equation (Model 5a). The positive MLE estimate of 0.142 (Model 5a) for the covariance σ_{SBIG4} and the negative estimate of –0.199 (Model 6a) for σ_{SNON} imply that an increase in the value for the unobservable error in the auditor selection equation (ε_{SEL}) is associated with an increase in the value of the unobservable component of Big Four fees (ε_{BIG4}) and a decrease in the value of unobservable component of non-Big Four fees (ε_{NON}), although the former estimate is insignificant at the 0.05 level. These results imply that the effect of unobservable auditee characteristics is to cause private companies to choose the most expensive auditor and directly contradict the results of Chaney et al. (2004). The results also imply non-Big Four auditees value each type of auditor differently from Big Four auditees: not only are they willing to pay more for non-Big Four auditors; they also place a lower value on the services of a Big Four auditor.[20]

The selection estimates of the Big Four premium are also dependent on the estimator used. As the results in Models 5b and 6b show, the standard Heckman two-step approach amplifies the MLE estimates. Although the λ coefficients have the same signs as their MLE counterparts, and are both significant at the 0.05 level in both equations, they are implausibly large (in absolute terms) at 0.446 for the Big Four and –0.509 for the non-Big Four auditees; thus the λ coefficient in the Big Four equation more than doubles (compared to the MLE). Further indications of model instability are provided by the insignificance of the intercept and *QUALIF* in the Big Four Heckman model (5b).

It has been acknowledged in prior research (Stolzenberg and Relles, 1997) that interpreting the magnitude of the λ coefficient is difficult due to the abstract nature of the variable itself. Extending the logic of our earlier decomposition to the Heckman results is informative in this context since it allows an assessment of the *effects* of the coefficient. To calculate the impact of selection bias on the Big Four premium by decomposing the observable and unobservable effects, we concentrate on the Big Four premium measured at the sample means of the Big Four auditees, i.e. the average effect of the treatment on the treated (*ATT*). Predicted fees paid by a Big Four client at the sample means are shown in Figure 7.

[18] The statistical insignificance of ln*SAL* in the auditor choice equation is not related to collinearity with *CHTA*. When *CHTA* was removed from Model 4, ln*SAL* remained statistically insignificant. In addition, when ln*SAL* was removed from Model 4, *CHTA* remained positive and statistically significant.

[19] The Wald statistic of 2668.40 ($p < 0.0001$) for Model 4b indicates the selection equation is well determined; the McFadden's R^2 is 0.204 and the model correctly classifies (cut-off point of 0.083 – representing the prior probability of selection into the Big Four) 77.52% and 70.81% of the Big Four and non-Big Four auditees respectively.

[20] Although *a priori* these findings may appear counterintuitive and may relate, as discussed below, to the lack of robustness of Heckman procedures in audit fee studies, they are not entirely implausible. As discussed earlier, there are numerous explanations (e.g. audit quality effects) for firms paying higher fees for Big Four audits; similarly, survey-based research by Marriott et al. (2007) finds that very small UK companies prefer non-Big Four auditors due to the more personal services and stronger relationships offered by smaller auditors. A further possibility is that the potential financial gains arising from switching auditor may not justify the associated costs.

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Figure 8
Estimates for Models 5a and 6a

$$\overline{\ln F_{BIG4}} = a + b + \sum_{k=1}^{K} b_k \overline{X_{k\,BIG4}} + g_{SBIG}\overline{\lambda_{BIG4,\,BIG4}}$$ For Big Four clients (actual) (13)

9.4364 = 9.2268 + 0.2096

And:

$$\overline{\ln F_{NON}} = a + \sum_{k=1}^{K} a_k \overline{X_{k\,BIG4}} + g_{SNON}\overline{\lambda_{BIG4,\,BIG4}}$$ For Big Four clients (counterfactual) (14)

8.8230 = 9.1174 – 0.2944

Table 4
Effects of changes in specification on MLE and standard Heckman two-step models

	Specification 1: full fee equation; full selection equation		*Specification 2: full fee equation; selection equation excludes CHTA*	
	MLE	*Heckman*	*MLE*	*Heckman*
Big Four λ coefficient	0.142	0.446	0.166	1.079
	(1.80)	(2.42)*	(2.31)*	(3.74)**
Non-Big Four λ coefficient	–0.199	–0.509	–0.204	–0.595
	(9.56)**	(8.39)**	(10.00)**	(9.22)**
Big Four conditional mean[Ψ]	9.436	9.436	9.436	9.436
Non-Big Four conditional mean	8.823	8.263	8.812	8.107
Difference (Big Four premium as *ATT*)	0.613	1.1729	0.624	1.329
Big Four unconditional mean	9.227	8.776	9.189	7.834
Non-Big Four unconditional mean	9.117	9.018	9.115	8.989
Difference (Big Four premium as *ATE*)	0.109	–0.242	0.074	–1.155
	(0.93)	(0.88)	(0.69)	(2.69)**
Big Four selection effect	0.210	0.660	0.246	1.602
Non-Big Four selection effect	–0.294	–0.755	–0.303	–0.883
Difference in selection effect	0.504	1.415	0.549	2.485
	(4.20)**	(4.92)**	(4.95)**	(5.66)**
Big Four R^2 (dep. var. = λ)[§]	0.986	0.986	0.992	0.992
Non-Big Four R^2 (dep. var. = λ)[§]	0.854	0.858	0.868	0.874

Notes:
The table reports maximum likelihood (MLE) and standard Heckman estimates for three specifications of the two-step correction model. Specification 1 corresponds to the MLE and Heckman models reported in Table 3. Absolute *t*-values are reported in parentheses.
**, * represent statistical significance at the 0.01 and 0.05 level, respectively.
[Ψ] The conditional mean is the predicted mean of audit fees conditional on the auditee choosing a Big Four or non-Big Four auditor, allowing for auditees' unobservable characteristics. The unconditional mean is the predicted mean of audit fees excluding the selection effects. The difference between the conditional and unconditional means is the selection effect.
[§] reports the R^2 for a regression of the selection term on the remaining variables in the audit fee (second stage) equation.

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Figure 9
Relationship between treatment effects

$$\overline{\ln F_{BIG4,BIG4}} - \overline{\ln F_{NON,BIG4}} = b + \Sigma_k (b_k - a_k)\overline{X_{kBIG4}} + g_{SBIG}\overline{\lambda_{BIG4,BIG4}} - g_{SNON}\overline{\lambda_{BIG4,BIG4}} \quad (15)$$

$(ATT) = (ATE) + \text{Estimate}[E(\varepsilon_{BIG4} \mid D = 1) - E(\varepsilon_{NON} \mid D = 1)]$

Difference in conditional means = Difference in unconditional means + Difference in unobservable effect

9.4364 – 8.8230	=	9.2268 – 9.1174	+	0.2096 + 0.2944
0.6134	=	0.1094	+	0.504

The counterfactual equation shows the predicted audit fees for a typical Big Four client which paid audit fees according to the non-Big Four model. Since the same regressor means are used to compute predicted audit fees, we have removed any potential differences due to the different characteristics (the explained differences) of the Big Four and non-Big Four auditees, with any remaining difference amounting to the Big Four premium (the unexplained differences). The two components of predicted audit fees estimate the separate effects of the observable regressors and the unobservables. The decomposition of the counterfactual audit fees ($\overline{\ln F_{NON}}$) comprises the predicted fees paid to a non-Big Four auditor by any firm with the same mean observable characteristics

$$(a + \sum_{k=1}^{K} a_k \overline{X_{kBIG4}})$$

plus the selection effect ($g_{SNON}\overline{\lambda_{BIG4,BIG4}}$) showing the predicted effect of unobservable characteristics. The first term represents the unconditional mean showing the predicted audit fees if the clients chose Big Four and non-Big Four auditors at random.[21] The predicted fees ($\overline{\ln F}$) incorporate the selection terms and are therefore referred to as the conditional means.[22] Hence, the predicted fee equations have the form: conditional mean = unconditional mean + selection effect. The estimates for Models 5a and 6a in Table 3 are shown in Figure 8.

These estimates, along with those for non-Big Four firms, are tabulated in Table 4. Thus the typical Big Four auditee actually paid fees in natural log form of 9.4364 (£12,537). By contrast it would have paid predicted fees as a non-Big Four auditee of 8.8230 in natural log form (£6,789), implying a Big Four premium of 0.6134 or 85%, which is much larger than that found in prior research. Our MLE results suggest that on average, Big Four auditees would have paid 9.2268 (£10,166) for the services of a Big Four auditor and 9.1174 (£9,113) for a non-Big Four auditor if their unobservable characteristics were ignored. However, in consequence of Big Four auditees' unobservable characteristics, an additional 0.2096 is paid for Big Four audit fees, and 0.2944 less for the services of a non-Big Four auditor.

The average treatment effect on the treated (*ATT*) is the difference in the conditional means of the audit fees paid by Big Four and non-Big Four auditees and represents the difference in fees only available to Big Four auditees. By contrast, the average treatment effect (*ATE*) shows the difference in fees available to any auditee. The relationship between the treatment effects is shown in Figure 9 (see Heckman et al., 2001).

Using the MLE parameters, the Big Four premium paid by Big Four auditees or *ATT* is 0.6134 (85%). The typical Big Four auditee paid 0.1094 (12%) more in fees based on their observable characteristics. This *ATE* (12%) is assumed to be freely substitutable, in that any non-Big Four auditee with the relevant characteristics who switched to a large auditor would incur this premium and vice versa. However the peculiar unobservable characteristics of Big Four auditees (as reflected in their value of λ) mean that they would pay an additional 0.504 in natural log terms for the services of a Big Four auditor, whereas the unobserved characteristics of non-Big Four auditees means they would be unwilling to pay this premium. Since the selection effects are individually significant, they should be included in the model. Although the *ATE* is not economically insubstantial at 12%, it is not significantly different from zero ($t = 0.93$; $p = 0.35$). By contrast the large selection effect is highly significant ($t = 4.20$; $p = 0.00$). According to our MLE results, therefore, firms with similar observable characteristics would pay higher fees if they used

[21] Note that with random selection, there would be no selection effect.

[22] They are conditional in the sense that they are estimates of the expected audit fees conditional on the firm employing either a Big Four or a non-Big Four auditor.

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

a Big Four auditor, but the difference is statistically insignificant. However, auditees differ greatly in their unobserved characteristics and these differences largely generate the Big Four premium. We emphasise that our results contrast sharply with Chaney et al. (2004), who report that unobservable factors make it *cheaper* for Big Four auditees to opt for Big Four auditors, rather than non-Big Four ones.

The Heckman two-step results in Table 4 (Specification 1) show that the typical Big Four auditee actually paid log fees of 9.436 (£12,537). By contrast it is predicted to have incurred fees as a non-Big Four auditee of 8.263 (£3,880) estimating the Big Four premium at an inconceivable 1.1729 or 223% (compared to 85% for the MLE estimates discussed above). Big Four auditees would have paid 8.776 (£6,478) for the services of a Big Four auditor and 9.018 (£8,250) for a non-Big Four auditor, if their unobservable characteristics were ignored. However, as a result of Big Four auditees' unobservable characteristics, an additional 0.660 (Big Four selection effect) is paid for the services of a Big Four auditor, and 0.755 less for the services of a non-Big Four auditor (non-Big Four selection effect). The effect of unobservables is to increase the Big Four fees (in £) by 93% ($e^{0.660}$=1.93). These estimates imply that Big Four auditees choose the cheaper auditor based on observed information, but the most expensive when unobservable characteristics are taken into account. Moreover, there are substantial differences between the MLE and Heckman estimates, which in itself, is indicative of model instability.

To assess the potential for parameter instability, Table 4 also examines the consequences of omitting the exclusion variable (*CHTA*) from the probit auditor selection equation. In this instance, changes in the MLE estimates are relatively minor, with the premium paid by the Big Four increasing moderately from 0.613 (85%) to 0.624 (87%). The difference in unconditional means (*ATE*) in the Heckman model, however, increases almost fivefold to –1.155 and is now statistically significant (at $p < 0.01$), although this is more than offset by the estimated impact of the unobservable selection difference (2.485). Comparing specifications therefore suggests the Heckman results are highly sensitive to model specification using the standard approach, although this is apparently less problematic when estimated using maximum likelihood.

The R^2 reported for both Big Four and non-Big Four equations in the bottom two rows of Table 4 for both models also testify to the high levels of multicollinearity when the λ variable is regressed on the other variables in the audit fee equation. In Specification 1, these are 0.986 and 0.854 (i.e. variance inflation factors of 71.4 and 6.8) for the Big Four and non-Big Four equations respectively, suggesting that even *with* an identifying variable (*CHTA*) multicollinearity may pose problems for the Big Four estimation. To summarise the results from Table 4, the ML estimator appears most efficient whereas the estimates provided by the standard Heckman two-step procedure (the only method used for dealing with selection bias in the auditing literature to date) are potentially seriously unstable.

In Table 5, we examine the sensitivity of the Heckman and ML estimates of the selection term (λ) by changing the samples and variables used to estimate the selection models; we also report estimates of the premium using a single stage OLS pooled model for comparison, together with the associated R^2 and sample size (n) for the various models.

The sample in row 1 is based on all Big Four and all non-Big Four auditees and is reported for comparative purposes. Although it was not our aim to replicate the Chaney et al. (2004) study, as a further robustness test, row 2 reports estimates for the current sample, using the specification employed by Chaney et al. in their study of UK private companies (and noting the absence of a valid identifying variable in the selection equation).[23] The Chaney et al. (2004) model estimated in our sample has a substantially higher explanatory power (R^2 = 0.68 for the pooled model) than reported in their study (0.57), although this is still substantially lower than for the model specification (row 1) in this study (0.78)[24,25]. Of more importance, however, is that the Heckman two-step estimates of the λ coefficients (row 2), are both highly significant ($p < 0.001$) and positive (1.850) and negative (–2.378) for the Big Four and non-Big Four models respectively. Similar findings are also evident from the MLE results. Hence, in contrast to Chaney et al. (2004), their models in our sample

[23] We are grateful to an anonymous reviewer for this suggestion. The variables are ln*TA*, *EXPSAL*, *LOND*, BUSY (all defined as in the current study), the ratio of exceptional and extraordinary items to total assets (*EXTA*), long-term debt to total assets, sales to total assets, the quick ratio, current assets to total assets, earnings before interest and taxes divided by total assets (*RTA*), *BUSY* × ln*TA* (*BTA*) and *RTA* × a loss indicator variable where unity = a company which made a loss in the prior period, zero otherwise. All variables were included in their audit fee model; but their selection model excluded *BUSY*, *BTA*, *LOND* and *EXTA*. There is a small fall in the sample size as a result of missing observations for the prior period loss indicator.

[24] Similar points also apply to the R^2 for separate Big Four and non-Big Four regressions.

[25] The lower explanatory power of the Chaney et al. (2004) models in their, relative to our, sample may partly relate to the larger sample (across all size ranges) employed in the current study and/or the rounding imprecision in their data noted earlier.

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Table 5
Effect of changes in sample and variables on OLS, Heckman two-step and MLE results

	Pooled OLS		*Heckman two-step*				*MLE*	
			Big Four		*Non-Big Four*		*Big Four*	*Non-Big Four*
Sample and variables	*BIG4* coefficient (*t*-value)	R^2 (*n*)	λ coefficient (*t*-value)	R^2 (*n*)	λ coefficient (*t*-value)	R^2 (*n*)	λ coefficient (*z*-value)	λ coefficient (*z*-value)
1. Current sample; current variables	0.270** (22.96)	0.78 (36,674)	0.446** (2.42)	0.80 (3,038)	–0.509** (8.39)	0.75 (33,636)	0.142 (1.80)	–0.199** (9.56)
2. Current sample; Chaney et al. (2004) variables	0.386** (27.10)	0.68 (36,607)	1.850** (3.71)	0.71 (3,026)	–2.378** (11.94)	0.67 (33,581)	0.277** (3.65)	–0.298** (19.67)
3. Sample firms with total assets ≥ £10,000; current variables	0.259** (21.99)	0.76 (34,351)	0.351* (2.38)	0.80 (3,023)	–0.313** (4.84)	0.73 (31,328)	0.140 (1.86)	–0.193** (8.29)
4. Thicker support region [φ]; current variables	0.280** (21.24)	0.72 (33,007)	0.075 (0.31)	0.69 (2,188)	–0.524** (5.08)	0.71 (30,189)	0.020 (0.16)	–0.179** (6.78)
5. Approximate matching sample[Ψ]; current variables	0.226** (18.91)	0.69 (24,505)	0.657 (1.81)	0.76 (2,734)	–0.025 (0.24)	0.66 (21,770)	0.065 (0.74)	–0.290** (–7.56)

Notes:
**, * represent statistical significance at the 0.01 and 0.05 level respectively; *t*-statistics for OLS models use robust standard errors and corrected standard errors for the Heckman two-step models.
All models labelled 'current variables' use the full model specification for the probit and audit fee equations in Table 3.
[φ] Thicker support region is represented by a sub-sample of firms where the top and bottom 5% of all sample firms are eliminated according to their probability of choosing a Big Four auditor.
[Ψ] Approximate matching sample is created by eliminating firms whose probability of selecting a Big Four auditor is greater (less) than the 95th (5th) percentile of the Big Four firms' probability of choosing a Big Four auditor.

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

produce the same signs on the λ coefficients for the Big Four and non-Big Four equations as those reported for the specifications employed in the current study (row 1) – although these coefficients are inconceivably large, being over four times larger than the respective coefficients of our Heckman two-step specifications. Furthermore, a contemporaneous working paper by Francis and Lennox (2008) reproduces the analysis of Chaney et al. (2004) using precisely the same variables and sample design (although for a later time period) confirms this instability. Inter alia, Francis and Lennox report an average positive λ coefficient (0.10) for the Chaney et al. Big Four auditor specification (as in the current study) and a positive one (0.21) for the non-Big Four specification. However, neither of the coefficients was statistically significant at conventional levels, suggesting no evidence of significant auditor selection bias. Finally, we note that using the Chaney et al. (2004) variables in our samples leads to a much higher OLS pooled estimate of the Big Four premium (47%) than that utilising the variables in this study (31%).

Since our auditee size cut-off point is low relative to previous studies, in row 3 we report parameter estimates for our previously reported models, but restricting the sample to firms with total assets exceeding £10,000.[26] The significance and signs of the λ coefficients are relatively stable – particularly the MLE parameters – although the R^2 of the pooled model declines (to 0.76) as does the estimated Big Four premium (to 29%).

In row 4 of Table 5, labelled the 'thicker support region', we exclude extreme observations (below the 5th and above the 95th percentiles) of the probability of selecting a Big Four auditor for the whole sample to provide estimates where there is more common support between the sub-samples. Row 5 excludes observations with a probability of selecting a Big Four auditor below the 5th, and above the 95th percentile for Big Four auditees and hence provides estimates where the Big Four and non-Big Four samples are more closely matched in terms of their selection probabilities. In row 4, the Big Four selection effect is now statistically insignificant for both the Heckman two-step and MLE parameters. In row 5, it is relatively large for the Heckman model but insignificant, as it is for the MLE model. These results underline the importance of comparable samples but at the possible cost of increasing multicollinearity. The non-linearity of the selection term is most noticeable for firms with high probabilities of selecting a Big Four auditor, so excluding these firms will tend to produce samples where a linear equation provides a more complete summary of the variation in λ. The converse applies for the non-Big Four firms and this is consistent with the Heckman two-step result in row 5, where large numbers of firms with high probabilities of choosing non-Big Four auditors are excluded.

In summary, unlike the standard single-stage pooled regression estimates of the Big Four premium, which Table 5 shows are always highly statistically significant in all models, this sensitivity analysis indicates that our original two-step results are not robust across all the sample and variable changes investigated. The sensitivity of the estimated selection effect to the omission of our identifying variable is consistent with extant econometric studies (above), and is of key importance to existing and future accounting studies which do not include a valid instrument when employing Heckman two-step procedures. The results in Table 5 also suggest that the positive selection effects for Big Four auditees may be driven by large companies and, with less certainty, the negative ones for non-Big Four auditees by small firms. There is evidence of serious instability in the coefficients, most noticeably varying by estimation technique. Nonetheless, the pattern of results gives no support to the hypothesis that Big Four auditees are choosing the lowest cost auditor after controlling for selection bias.

4.3. Matching results

Because of the sensitivity of the Heckman model demonstrated above, this section reports the results of the Big Four premium matching analysis. Under the conditional independence assumption, this produces estimates of the average treatment effect on the treated (*ATT*) not prone to the model specification and identification problems discussed above. We employ two matching methods: propensity score matching and a pre-process matching analysis combined with OLS regressions.

Propensity score matching

As discussed above, recent developments in the statistics and econometrics literature have suggested propensity score matching as an additional or alternative approach to two-step Heckman procedures. Since the seminal paper of Rosenbaum and Rubin (1983), propensity score matching has received considerable attention as a means of estimating causal treatment effects. In our analysis, Big Four auditees are matched to non-Big Four auditees on the basis of the predicted probability of employing a Big Four auditor – with the propensity scores (predicted probabilities) being derived from the probit selection equation (Model 4, Table

[26] We are grateful to the anonymous reviewers for suggesting this extension of our analysis. We also conducted this analysis for firms above and below the median and restricting the sample to firms with total assets over £1,000,000 and obtained further evidence of instability.

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

3), which includes all explanatory variables listed in Table 1.[27] Following matching, the average audit fees paid by the matched sub-samples are compared, to assess whether a significant premium is evident.[28]

The most popular propensity score matching method is nearest neighbour matching. In our study, this entails matching each Big Four auditee to the non-Big Four counterpart with the propensity score closest in value to that of the Big Four auditee. Nearest neighbour matching can be implemented with or without a 'calliper', where the calliper represents the maximum (absolute) difference between the propensity score of the nearest neighbour matched observations. A tighter calliper results in more closely matched observations but reduces the sample size, i.e. it only selects observations that can be matched within the minimum distance imposed by the calliper. When employing this method, researchers face a choice of whether to use replacement observations, i.e. permitting the use of non-Big Four (non-treated) auditees for matching with their Big Four counterparts more than once. This can be important in the nearest neighbour (without calliper) approach, since very large Big Four clients may have a limited number of counterparts in the non-Big Four sample; hence, excluding replacement can result in relatively large differences in propensity scores between the matched observations.

We therefore report in Table 6 (Panel A) results based on four different matching approaches to examine the robustness of our results and to illustrate the differences between the various methods,[29] each of which produces an equal number of matched Big Four and non-Big Four auditees. Panel A of Table 6 shows that the Big Four premium (the difference in the means of ln*AFEE* of the Big Four and non-Big Four sub-samples) is statistically significant at $p < 0.01$ under each type of matching, ranging from 0.2531 (28.8%) with a calliper of 0.001 (column 4) to 0.3082 (36.1%) under the nearest neighbour method with no replacement in column 3.[30] Moreover, as the statistics in column 5 demonstrate, even when Big Four and non-Big Four companies are very closely matched (with a maximum absolute difference in propensity scores of only 0.0001), the premium (0.2613 or 29.9%) remains robust and is within the range of premiums reported in prior research and in our earlier analysis based on OLS.

Although the results in column 5 demonstrate that the samples are very closely matched on their observed characteristics based on the composite propensity score, it is also important to examine how closely matched the Big Four auditees are to their non-Big Four counterparts in respect of the individual covariates. Panel B of Table 6 reveals that the two samples are also very similar in respect of each of the individual characteristics (variables) in the auditor choice and fee equations (Model 4). Indeed, whilst there are substantial (and significant) differences between the characteristics of the Big Four and non-Big Four auditees before matching (see Table 2), after matching (using the finest calliper of 0.0001) the differences in the means become small and are all statistically insignificant.[31]

Since our sample includes a large number of smaller auditees (due to our relatively low cut-off of £1,000 for total assets), we also conducted additional propensity score matching analysis on samples restricted first to companies with total assets (*TA*) in excess of £10,000 and second to those with *TA* over £1m. Using the 0.0001 calliper, we found that for the first sample (i.e. where *TA* > £10,000), the premium was estimated at 0.2311 (26.0%) and for the second sample (where *TA* > £1m) at 0.1753 (19.2%). Both estimates were statistically significant and even although the latter is somewhat lower than the estimated premium for our main sample, it is within the range reported by Moizer (1997). Furthermore, using the wider calliper of 0.001 on both these size-restricted samples and on two separate sub-samples of firms with *TA* above and below the median (for the full sample) yielded very similar results. Finally, we estimated the models without the instrument (*CHTA*) and our findings were unchanged.[32]

On the basis of matched samples that are very similar in terms of their observed characteristics, therefore, we find strong evidence of a Big Four premium of a similar magnitude to that found in studies employing OLS. Unlike those provided by the Heckman approach, these estimates are not sensitive to model specification and do not impose assumptions of linearity. Moreover, variation in

[27] The conditional mean independence assumption is that the choice of regime (Big Four auditee or non-Big Four auditee) is not dependent on the regime once the matching variables (*Z*) are taken into account. This means, in practice, that the values of *Z* should not depend on the type of regime. We therefore use all the regressors in the fees equation as matching instruments (*Z*) and make the reasonable assumption that all the measured characteristics are pre-determined before the choice of auditor is made.

[28] Leuven and Sianesi (2003) provide details of the propensity score matching module (psmatch2) for use with Stata statistical software.

[29] The analysis in Table 6 is based on differences in the log audit fees to allow comparison with previous findings. We also conducted this analysis using untransformed audit fees and obtained very similar results.

[30] The results based on the 0.001 and 0.0001 callipers in columns 4 and 5 of Table 6 are conducted without replacement, though they are virtually unchanged when we do allow replacement, with a mean difference in log audit fees between Big Four and non-Big Four auditees of 0.2566 and 0.2639 for the 0.001 and 0.0001 callipers respectively (both estimates were statistically significant).

[31] We find similar (unreported) results using the 0.001 calliper, where the only variable approaching statistical significance is *SQSUBS* (difference in means of –0.06; $p = 0.066$).

[32] These additional results are unreported for brevity but are available from the authors upon request.

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Table 6
Propensity score matched results[†]

Panel A: Alternative propensity score matching estimators

	Nearest neighbour (with replacement)	*Nearest neighbour (no replacement)*	*Calliper of 0.001*[Ψ] *(no replacement)*	*Calliper of 0.0001 (no replacement)*
Mean difference in ln*AFEE*	0.2642	0.3082	0.2531	0.2613
Big Four premium[‡]	30.2%	36.1%	28.8%	29.9%
z-statistic	8.23**	15.58**	12.54**	10.91**
N[§]	6076	6076	5586	4814
Mean difference in p-score[††]	0.0003	0.2393	0.0001	0.0000
Min. difference in p-score	0.0000	0.0000	0.0000	0.0000
Max. difference in p-score	0.3960	0.6848	0.0010	0.0001

Panel B: Effects of matching (calliper 0.0001) on covariate and propensity score means[φ]

Variable	*Big Four pre-match*	*Non-Big Four pre-match*	*Big Four post-match*	*Non-Big Four post-match*
P-SCORE[τ]	0.2246	0.0700	0.1582	0.1582
ln*SAL*	15.666	13.385	15.128	15.149
ln*TA*	15.488	12.860	14.906	14.884
CHTA	5.2616	0.4891	1.3583	1.2821
SQSUBS	1.1714	0.2999	0.7827	0.8236
EXPSAL	0.7871	0.0235	0.0571	0.0553
QUALIF	0.0375	0.0315	0.0420	0.0403
PBAL	0.1208	0.0380	0.0835	0.0827
CONLIAB	0.2749	0.0967	0.2194	0.2231
EXITEM	0.5662	0.3206	0.5131	0.5172
TLTA	0.8433	0.7731	0.7817	0.7493
RTA	0.0087	0.2279	0.0206	0.0218
LOND	0.2094	0.3426	0.2185	0.2019
BUSY	0.5583	0.4399	0.5231	0.5330

Notes
[†] The probit selection model from which propensity scores (p-scores) are derived is reported in Table 3 (Model 4).
[††] p-scores are propensity scores.
[‡] Results are based on bootstrapped standard errors (50 replications). The premium is the difference between the mean ln*AFEE* for Big Four auditors and the mean ln*AFEE* for matched Big Four counterparts.
[§] Note that each method results in an equal number of matched Big Four and non-Big Four auditees.
[Ψ] The calliper is the maximum permitted absolute difference in propensity score between matched observations.
** represents statistical significance at the 0.01 level.
[φ] None of the means in Panel B differed significantly after matching.
[τ] Note that the mean (p) of the predicted probabilities derived from the probit model for the whole sample always equals the prior probability of selection into the unity value of the binary dependent variable; in our case $p = 0.083$ (the proportion of Big Four auditees in the sample).

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Table 7
Pre-processed portfolio matched regression results

	Mean	*Std. dev.*	*Median*	*Min*	*Max*
BIG4 coefficient	0.2724	0.0114	0.2724	0.2369	0.3091
Big Four premium	31.3%	–	31.3%	26.7%	36.2%
BIG4 *t*-statistic	13.73	0.5652	13.73	11.98	15.51
F-Value	832.58	30.15	833.24	717.57	926.04
Adjusted R^2	0.7679	0.0040	0.7679	0.7544	0.7792

Notes
† This table reports the distribution of the *BIG4* coefficient estimate from 2,000 iterations, where each iteration involves a regression (see Model 1 in Table 3) using a total sample of 3,656 companies (i.e. 1,828 Big Four and 1,828 non-Big Four auditees), together with associated *t*-statistics and model *F* and R^2. The total sample of 36,674 companies is first divided on the basis of sales (40 quantiles), exports to sales (11 quantiles), return on total assets (40 quantiles) and the number of subsidiaries (10 quantiles). Each Big Four auditee is then matched to a non-Big Four counterpart belonging to the same quantile for each of the four variables. The number of quantiles for the ratio of exports to sales and number of subsidiaries differs due to a large number of zero values for each variable. The *t*-statistic in each iteration uses robust standard errors.

the matching approach (e.g. changing callipers and allowing replacement observations) indicates that, unlike the Heckman approach, the results are not sensitive (in terms of a significant premium) to changes in samples. It should be emphasised, however, that the additional robustness, and the fact that no assumptions are made about functional form in assessing the premium, come at a cost: any unobservable differences between Big Four and non-Big Four clients which systematically influence both auditor choice and audit fees are effectively assumed to be randomly distributed across the samples of Big Four and non-Big Four clients.

Pre-processed OLS results

With our pre-processed analysis, using Stata statistical software, we first partition our sample of 36,674 companies into quantiles on the basis of their actual size, risk and complexity, as these factors have been found to be particularly important determinants of both audit fees and auditor selection (e.g. see Simunic and Stein, 1996; Chaney et al., 2004). We created 40 equally sized quantiles based on sales (*SAL*), 40 equally sized quantiles based on return on total assets (*RTA*), 10 quantiles based on the number of subsidiaries (*SUBS*) and 11 quantiles based on the ratio of exports to sales (*EXPSAL*). We then partition the above quantile samples into companies audited by Big Four and non-Big Four auditors and matched them, so that each individual Big Four auditee had an individual non-Big Four counterpart with concurrent membership of the same size quantiles for *SAL*, *RTA*, *SUBS* and *EXPSAL*. Hence observations are matched where they exhibit (jointly) similar size, risk and complexity characteristics. Note that this pre-processing method can be a more demanding process than propensity score matching since the former matches on the basis of actual values for the four control variables *simultaneously*, rather than on one composite score; that is, each Big Four firm has a non-Big Four counterpart with similar observed size *and* risk *and* complexity characteristics.

A dilemma associated with this matching process is that many Big Four clients have a number of non-Big Four counterparts of similar size, risk and complexity. In order to circumvent this problem, we randomly selected (with replacement) one match for each Big Four auditee. We then combined the non-Big Four auditee sample and the Big Four auditee sample and re-estimated our standard regression equation (Model 1 in Table 3). We repeated this process 2,000 times and obtained a distribution of *BIG4* regression coefficients and their associated robust (White's corrected) *t*-statistics (White, 1980). Each iteration involved samples of 1,828 Big Four auditees and 1,828 matched non-Big Four auditee counterparts (i.e. a total sample in each regression of 3,656). Additional analysis (unreported but available on request) showed that the two (Big Four and non-Big Four) auditee samples were very closely matched on the four matching variables with, on average, all differences being highly statistically insignificant.[33]

[33] For each iteration, we collected the *p*-value and *t*-statistic for mean differences between the Big Four and non-Big Four samples for the four matching variables and out of the 2,000 iterations, there were no significant differences (at $p < 0.05$) in the variables on which we matched. More specifically, the range and mean for the absolute *t*-statistics, respectively, of the four variables were 0.31–0.56 and mean of 0.44 for ln*SAL*; 0.17–1.90 and mean of 1.07 for *RTA*; 0.18–0.51 and mean of 0.34 for *SQSUBS*; and 0.02–0.04 with a mean of 0.04 for *EXPSAL*. We also repeated this procedure by controlling for both sales and total assets (together with *SUBS*, *EXPSAL* and *RTA*) and obtained similar results (though inevitably on a smaller sample).

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Descriptive statistics from the 2,000 regressions for the *BIG4* coefficient and robust *t*-statistics are reported in Table 7. The table shows that in every case, the *BIG4* coefficient was statistically significant and positive, ranging from 0.2724 (26.7%) to 0.3091 (36.2%), with the mean and median taking the same value of 0.2724 (31.3%). The distribution of *t*-statistics (which are based on White's corrected standard errors) reveals that the *BIG4* coefficient is consistently significant at $p < 0.01$. It is also interesting to note that the range of coefficients implies a premium between 27% and 36%, which is broadly in line with that (16–37%) found in prior literature and of a similar magnitude to our propensity score estimates. Hence, the Big Four premium is persistent after matching on key auditee attributes (size, risk and complexity) and controlling for any remaining confounding influences via OLS regression.

5. Conclusions

Since the seminal paper by Simunic (1980), a large number of studies have predicted and found that large auditors have commanded a premium for their services, possibly due to superior audit quality, 'deep pockets' and other reputational effects. Important innovations in the literature by (inter alia) Ireland and Lennox (2002) and Chaney et al. (2004) challenged findings based on OLS regressions. The latter paper overturned much of the prior research by stating that, given their firm specific characteristics, private UK companies that chose a Big Five auditor would have paid more had they chosen a non-Big Five auditor, thus leading to the conclusion that the large auditor premium does not exist and that the audit market is properly organised (*ibid*.: 70). Although this is an economically persuasive conclusion, the econometrics literature suggests that selection effects estimated using the Heckman procedure may be highly sensitive to model specification and collinearity; consequently, standard OLS regression can produce more accurate estimates than their two-stage counterparts (Hartman, 1991; Stolzenberg and Relles, 1997).

Applying the Oaxaca-Blinder decomposition to OLS regression results, we find that the majority of the Big Four premium is attributable to large differences in the characteristics of Big Four and non-Big Four auditees. Using this approach, we estimate the Big Four premium at 29–31%, using the estimates of linear equations for samples with markedly different characteristics. An important question, however, is whether the premium correctly reflects what the change in fees would be if auditees changed auditor type. The OB decomposition assumes each auditee can change auditor (unconstrained) to the other type and pay the corresponding counterfactual fees.[34] Our results based on the Heckman correction (which allows for unobservable differences between Big Four and non-Big Four auditees) confirm prior reports in other areas of social science research that this procedure does not represent a panacea for estimating selection effects. The standard Heckman two-step method is highly sensitive to model specification, collinearity and to sample composition. The results also appear highly dependent on the estimation technique adopted. A comparison of the standard Heckman two-step and maximum likelihood estimators revealed large differences in the estimate of the selection effect.

Our findings relating to the impact of correctly identifying the Heckman model with an appropriate instrument are of key importance, since the difficulty in obtaining such an exclusion variable is often a major barrier to valid implementation of the method (see, e.g. Bryson et al., 2002); and ideally, the instrument should *not* be strongly correlated with the remaining selection model regressors. Furthermore, although we present evidence of a significant selection effect, this does not support the interpretation of cost minimisation and does suggest that Big Four auditors receive higher audit fees. The potentially critical finding of Chaney et al. (2004) that private firms select the type of auditor that provides the cheapest service once unobservable auditee characteristics are taken into account is therefore unsupported by our analysis of private UK independent firms.

Whilst two-stage estimators allow for potential unobserved characteristics, our results demonstrate that the advantages of these methods are unproven, and should be traded off against their sensitivity. Three different studies of UK private companies have produced three different sets of results relating to the impact of unobservable variables on premiums using the Heckman estimator, but all reported significant premiums using standard models. Although it is possible (but unlikely) that unobservable factors vary systematically between these studies, further research seems warranted into the causes of this variation (e.g. whether unobservables are non-linear in form, model identification and explanatory variables utilised), and the nature of any such characteristics.

When we employ a more stable matching approach to estimate the premium by comparing the audit fees for Big Four and non-Big Four auditees of a similar degree of size, risk and complexity, we find a persistent premium of a magnitude in line with that found in prior single-stage OLS audit fee studies. The propensity score matching and preprocessed matched regression estimates of the Big

[34] We noted in the paper, for example, that this is highly improbable in respect of the largest auditees (e.g. FTSE 100 companies), which are *de facto* constrained to Big Four auditors. Linear extrapolation outside the common support region in such cases may lead to inaccurate premium estimates; and here propensity score and pre-processed portfolio matching approaches may provide more accurate estimates.

Four premium are around 30%, with such matching estimators being highly robust to changes in model specification, and uninfluenced by assumptions of functional form. Although extant analytical research suggests unobservable factors such as insider knowledge of future cash flows may be important in determining auditor choice, there is disagreement on the magnitude and direction of these effects (cf. Titman and Trueman, 1986; Datar et al., 1991). Further research is warranted on the identification and implementation of empirical proxies for these characteristics (including variables capturing the presence and efficiency of the internal audit function and financial information and control systems), and the inclusion of such variables in more robust matching analyses. For example, Collis et al. (2004) find, using questionnaire data, that demand for voluntary audit of small companies is related to the perceived improvements in internal controls and in the quality of information provided (e.g. to banks). Such methods may provide insights into differential demand for Big Four and non-Big Four audits although may themselves be subject to measurement bias. Taken together with the findings of previous research that large auditors produce higher quality audits (e.g. Blokdijk et al., 2006), and with the findings of a contemporaneous study by Francis and Lennox (2008), our results suggest that the Big Four premium persists and that Heckman-based research that the premium vanishes once selection is allowed for should be treated with caution.

Appendix

(i) Oaxaca-Blinder decomposition

Standard pooled OLS regression using a Big Four dummy indicator variable in a single equation assumes that all slope coefficients for the Big Four and non-Big Four are identical (i.e.using the notation from equations (1) and (2) in the text, $\alpha_k = \beta_k$, all k). The Oaxaca-Blinder decomposition is based on separate regressions for Big Four and non-Big Four clients and emphasises that the observed actual difference in audit fees can partly be attributed to the different characteristics of the two types of auditees and partly to any Big Four premium. Hence, it expresses the difference in the means of audit fees as:

$$\overline{\ln F_{BIG4}} - \overline{\ln F_{NON}} = \sum_{k=1}^{K} a_k(\overline{X_{kBIG4}} - \overline{X_{kNON}}) + b + \sum_{k=1}^{K}(b_k - a_k)\overline{X_{k\,BIG4}} = EXPLAINED_{BIG4} + P_{BIG4} \quad \text{(A1)}$$

$$\overline{\ln F_{BIG4}} - \overline{\ln F_{NON}} = \sum_{k=1}^{K} b_k(\overline{X_{kBIG4}} - \overline{X_{kNON}}) + b + \sum_{k=1}^{K}(b_k - a_k)\overline{X_{k\,NON}} = EXPLAINED_{NON} + P_{NON} \quad \text{(A2)}$$

(ii) Heckman selection model

The Heckman procedure involves estimating the selection term λ by modelling the auditor choice process in the first step. Here, each company has an unobserved propensity (D^*) to choose a Big Four auditor. D^* is a linear function of M regressors (Z_m $m = 1,.., M$) and other unobservable characteristics (ε_{SEL}). The model can therefore be represented as follows:

$$D^* = \delta + \sum_{m=1}^{M} \delta_m Z_m + \varepsilon_{SEL} \quad \text{Auditor choice equation} \quad \text{(A3)}$$

$$\ln F_{BIG4} = \alpha + \beta + \sum_{k=1}^{K} \beta_k X_{k\,BIG4} + \varepsilon_{BIG4} \quad \text{For Big Four auditees} \quad \text{(A4)}$$

$$\ln F_{NON} = \alpha + \beta + \sum_{k=1}^{K} \beta_k X_{k\,NON} + \varepsilon_{NON} \quad \text{For non-Big Four auditees} \quad \text{(A5)}$$

If $D^* > 0$, $D = 1$ and we observe $\ln F = \ln F_{BIG4}$. Otherwise $D^* \leq 0$, $D = 0$ and $\ln F = \ln F_{NON}$. The model assumes that the errors of the selection and fee equations are jointly normal with zero means, constant variances and covariances: $E(\varepsilon_{SEL}\varepsilon_{NON}) = \sigma_{SNON}$ and $E(\varepsilon_{SEL}\varepsilon_{BIG4}) = \sigma_{SBIG4}$. The implied correlation between the unobservable factors determining the choice of auditor type and audit fees, together with the assumption of normality, enable the estimation of this model.[35] The importance of the normality assumption is discussed by Greene (2003: 789)

[35] For instance, companies more likely to employ Big Four auditors (i.e. have 'large' ε_{SEL}) given their observable characteristics (Z) are likely to value unobservable aspects of Big Four auditors' services more highly (i.e. have 'large' ε_{BIG4}).

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

and by van der Klaauw and Koning (2003); with the latter reporting that coefficient estimates do not appear sensitive to departures from normality. The model may be estimated directly by maximum likelihood but it is more common to employ the Heckman two-step method. This involves the estimation of an augmented regression equation with necessary adjustments to the formulae for the standard errors:

$$\ln F_{BIG4} = \alpha + \beta + \sum_{k=1}^{K} \beta_k X_{k\,BIG4} + \sigma_{SBIG4}\lambda_{BIG4} + \upsilon \qquad \text{For Big Four auditees} \qquad \text{(A6)}$$

$$\ln F_{NON} = \alpha + \sum_{k=1}^{K} \alpha_k X_{k\,NON} - \sigma_{SNON}\lambda_{NON} + \upsilon \qquad \text{For non-Big Four auditees} \qquad \text{(A7)}$$

$$\text{where} \quad \lambda_{BIG4} = \frac{\phi(\sum_{m=1}^{M}\delta_m Z_m)}{\Phi(\sum_{m=1}^{M}\delta_m Z_m)} \quad \text{and} \quad \lambda_{NON} = \frac{\phi(-\sum_{m=1}^{M}\delta_m Z_m)}{\Phi(-\sum_{m=1}^{M}\delta_m Z_m)} \qquad \text{(A8)}$$

and ϕ is the normal density function and Φ the normal distribution function. Note that the fee equation for non-Big Four auditees is estimated with selection into non-Big Four (i.e. the dependent variable for the probit is $ND = 1$ if the firm is a non-Big Four auditee). The coefficient of the selection term in this estimation is therefore the covariance between the error in the selection equation determining whether $ND = 1$ and ε_{NON}, i.e. an estimate of $-\sigma_{SNON}$. In the interests of comparability between the two equations, all results in the paper for the non-Big Four fee equation report estimates of σ_{SNON}.

(iii) Matching estimators

Each Big Four auditee can be defined in terms of the values of some observable variables ($\boldsymbol{Z} \equiv \{Z_1,..,Z_M\}$). Matching uses $\boldsymbol{Z}$ to produce a comparable non-Big Four auditee for each Big Four auditee. Let M be the set of N_M matched pairs of firms. The estimated treatment effect for each matched Big Four auditee is:

$$D(Z_i) = \ln F_{BIG4}(Z_i) - \ln F_{NON}(Z_i) \qquad i \in M \qquad \text{(A9)}$$

The estimated Big Four premium (Δ), or the treatment effect, is the sample mean of these differences across all values of Z in M or the difference in the sample means. Hence:

$$\Delta = \left[\frac{1}{N_M}\sum_{i\in M} \ln F_{BIG4}(\boldsymbol{Z}_i)\right] - \left[\frac{1}{N_M}\sum_{i\in M} \ln F_{NON}(\boldsymbol{Z}_i)\right] \qquad \text{(A10)}$$

$$\Delta = \overline{\ln F_{BIG4M}} - \overline{\ln F_{NONM}} \qquad \text{(A11)}$$

where the subscript M indicates that the mean is for companies in the matched sample.

Matching relies heavily on two assumptions: the conditional independence assumption (CIA) and common support. The CIA requires the value of audit fees to be independent of auditor type given the values of the observable variables.[36] To illustrate, assume that auditee size is the only determinant of audit fees. Consider a simple comparison of the mean audit fees paid by all Big Four auditees with those paid by all non-Big Four auditees, thus treating auditee size as an unobservable variable. This only makes sense if auditee size is not a determinant of auditor choice. However, if we compare samples of Big Four and non-Big Four auditees with the same size distribution we can ignore the effect of auditee size. CIA states formally that any unobservable variation in fees after adjusting for auditee size has the same random distribution for each type of auditee. Such adjustments are only possible if Big Four and non-Big Four auditees exist of comparable size.

The common support assumption merely states that such comparable auditees exist. For example, if all very large auditees were only audited by the Big Four, then it would be impossible to find comparable non-Big Four auditees. If the Big Four premium is regarded as payable by any auditee, matching assumptions imply that auditees are freely able to switch auditor and incur the corresponding counterfactual fees. It is therefore assumed that any systematic effect of the choice of auditor (D) on audit fees can be entirely explained in terms of some observable variables ($\boldsymbol{Z}$). In practice, $\boldsymbol{Z}$ is interpreted as the set of determinants of the auditor choice decision. Consider the following semi-parametric matching model (with the selection equation repeated for reference):

[36] More formally: $\ln F_{BIG4}, \ln F_{NON} \perp D \mid \boldsymbol{Z}$

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

$$D^* = \delta + \sum_{m=1}^{M} \delta_m Z_m + \varepsilon_{SEL} \qquad \text{Auditor choice equation} \qquad \text{(A12)}$$

$$\ln F_{BIG4} = \mu_{BIG4}(\mathbf{Z}) + \varepsilon_{BIG4} \qquad \text{For Big Four auditees} \qquad \text{(A13)}$$

$$\ln F_{NON} = \mu_{NON}(\mathbf{Z}) + \varepsilon_{NON} \qquad \text{For non-Big Four auditees} \qquad \text{(A14)}$$

If $D^* > 0$, $D = 1$ and we observe $\ln F = \ln F_{BIG4}$. Otherwise $D^* \leq 0$, $D = 0$ and $\ln F = \ln F_{NON}$. The audit fee equations above have additive errors but may be nonlinear in the conditional mean values (μ_{BIG4} and μ_{NON}). These means may depend on other regressors but the CIA means that the only relevant determinants are contained in ($\mathbf{Z}$). It also means that differences in unobservable auditee characteristics conditional on $\mathbf{Z}$ are random. Matching therefore provides an elegant method of eliminating explained differences. Since it compares samples of auditees with the same characteristics, it is unnecessary to specify the functional form for $\mu_{BIG4}(\mathbf{Z})$ and $\mu_{NON}(\mathbf{Z})$ and to rely on linear projections to produce counterfactuals – an important advantage where a non-linear relationship whose form is often unknown may exist between the dependent variable and the regressors. In practical applications, propensity score matching overcomes the 'curse of dimensionality' problem (which is encountered where $\mathbf{Z}$ contains numerous variables, resulting in a small number of suitable observations for matching purposes) by matching only on one variable (the propensity score), defined as:

$$p(\mathbf{Z}) \equiv \Pr(D = 1 \mid \mathbf{Z}) \qquad \text{(A15)}$$

The estimated premium for each matched Big Four auditee is defined as:

$$\Delta(\mathbf{Z}_i) = \ln F_{BIG4}(p(\mathbf{Z}_i)) - \ln F_{NON}(p(\mathbf{Z}_i)) \qquad i \in \Omega \qquad \text{(A16)}$$

where Ω is the set of N_Ω matched pairs of firms. The estimated Big Four premium (Δ) is the sample mean of these differences across all values of $p(\mathbf{Z})$ in Ω, or the difference in the sample means:

$$\Delta = \left[\frac{1}{N_\Omega} \sum_{i \in \Omega} \ln F_{BIG4}(p(\mathbf{Z}_i))\right] - \left[\frac{1}{N_\Omega} \sum_{i \in \Omega} \ln F_{NON}(p(\mathbf{Z}_i))\right] \qquad \text{(A17)}$$

References

Beatty, R.P. (1993). 'The economic determinants of auditor compensation in the initial public offerings market', *Journal of Accounting Research*, 31(2): 294–302.

Black, D. and 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.

Blinder, A. (1973). 'Wage discrimination: reduced form and structural variables'. *Journal of Human Resources*, 8: 436–455.

Blokdijk, H., Drieenhuizen, F., Simunic, D. and Stein, M. (2006). 'An analysis of cross-sectional differences in big and non-big public accounting firms' audit programs'. *Auditing: A Journal of Practice and Theory*, 25(1): 27–48.

Bryson, A., Dorsett, R. and Purdon, S. (2002). The use of propensity score matching in the evaluation of active labour market policies, Working Paper, No. 4. HMSO, London.

Chan, P., Ezzamel, M. and Gwilliam, D. (1993). 'Determinants of audit fees for quoted UK companies'. *Journal of Business Finance & Accounting*, 20(6): 765–786.

Chaney, P.K., Jeter, D.C. and Shivakumar, L. (2004). 'Self-selection of auditors and audit pricing in private firms'. *Accounting Review*, 79(1): 51–72.

Chaney, P.K., Jeter, D.C. and Shivakumar, L. (2005). 'Self-selection of auditors and size nonlinearities in auditing'. Working paper, Vanderbilt University.

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

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.

Copley, P.A. and Douthett, E.B. (2002). 'The association between auditor choice, ownership retained, and earnings disclosure by firms making initial public offerings'. *Contemporary Accounting Research*, 19(1): 49–75.

Craswell, A., Francis, J. and Taylor, S. (1995). 'Auditor brand name reputations and industry specializations'. *Journal of Accounting and Economics*, 20: 297–322.

Datar, S.M., Feltham, G.A. and Hughes, J.S. (1991). 'The role of audits and audit quality in valuing new issues'. *Journal of Accounting and Economics*, 14: 3–49.

Doyle, J.T., Ge, W. and McVay, S. (2007). Accruals quality and internal control over financial reporting'. *The Accounting Review*, 82(5): 1141–1170.

Ezzamel, M., Gwilliam, D. and Holland, K. (1996). 'Some empirical evidence from publicly quoted companies on the relationship between the pricing of audit and non-audit services'. *Accounting and Business Research*, 27(1): 3–16.

Firth, M. (1997). 'The provision of non-audit services and the pricing of audit fees'. *Journal of Business Finance & Accounting*, 24(3): 511–525.

Francis, J. and Lennox, C. (2008). 'Selection models in accounting research', HKUST Working paper, available at: http://ssrn.com/abstract=1120796.

Francis, J., Maydew, E.L. and Sparks, H.C. (1999). 'The role of big 6 auditors in the credible reporting of accru-

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

Downloaded by [Universitas Dian Nuswantoro], [Ririh Dian Pratiwi SE Msi] at 20:11 29 September 2013

als'. *Auditing: A Journal of Practice and Theory*, 18: 17–34.

Francis, J. and Simon, D.T. (1987). 'A test of audit pricing in the small-client segment of the U.S. audit market'. *Accounting Review*, 62(1): 145–157.

Giles, M. (2003). 'Correcting for selectivity bias in the estimation of road crash costs'. *Applied Economics*, 35(11): 1291–1301.

Giroux, G. and Jones, R. (2007). 'Investigating the audit fee structure of local authorities in England and Wales'. *Accounting and Business Research*, 37(1): 21–37.

Greene, W.H. (2003). *Econometric Analysis*. New York, NY: Prentice Hall.

Hamilton, J., Li, Y. and Stokes, D. (2005). 'Listed company audit self-selection bias and audit fee premiums'. Working paper, University of Technology Sydney.

Hartman, R. (1991). 'A Monte Carlo analysis of alternative estimators in models involving selectivity'. *Journal of Business and Economic Statistics*, 9: 41–49.

Hay, D.C., Knechel, W.R. and Wong, N. (2006). 'Audit fees: a meta analysis of the effect of supply and demand attributes'. *Contemporary Accounting Research*, 23(1): 141–191.

Heckman J., Tobias, J.L. and Vytlacil, E. (2001). 'Four parameters of interest in the evaluation of social programs'. *Southern Economic Journal*, 68(2): 210–223.

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

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

Keasey, K. and Watson, R. (1991). 'An agency perspective of auditor change in small firms'. *Journal of Small Firm Finance*, 1(1): 45–59.

Lennox, C. (1999). 'Are large auditors more accurate than small auditors?'. *Accounting and Business Research*, 29(3): 217–228.

Leung, S.F. and Yu, S. (2000). 'Collinearity and two-step estimation of sample selection models: problems, origins, and remedies'. *Computational Economics*, 15: 173–199.

Leuven, E. and Sianesi, B. (2003). 'PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing'. http://ideas.repec.org/c/boc/bocode/s432001.html.

Li, K. and Prabhala, N.R. (2007). 'Self-selection models in corporate finance'. In B.E. Eckbo (ed.) *Handbook of Corporate Finance*, Vol. 1. Amsterdam: Elsevier.

Little, R. (1985). 'A note about models for selectivity bias'. *Econometrica*, 53: 1469–1474.

Little, R. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York, NY: John Wiley & Sons.

Marriott, N., Marriott, P., Collis, J. and Son, D. (2007). 'The accountant and the provision of financial advice to UK smaller companies'. Working paper, University of Winchester.

McMeeking, K., Peasnell, K. and Pope, P.F. (2006). 'The determinants of the UK big firms premium'. *Accounting and Business Research*, 36(3): 207–231.

McMeeking, K., Peasnell, K. and Pope, P.F. (2007). 'The effect of audit firm mergers on audit pricing in the UK'. *Accounting and Business Research*, 37(4): 301–319.

Moizer, P. (1997). 'Auditor reputation: the international empirical evidence'. *International Journal of Auditing*, 1(1): 61–74.

Neumayer, E. (2003). 'Do human rights matter in bilateral aid allocation? A quantitative analysis of 21 donor countries'. *Social Science Quarterly*, 84(3): 650–666.

Oaxaca, R. (1973). 'Male-female wage differentials in urban labor markets'. *International Economic Review*, 14: 693–709.

Oxera (2006). *Competition and Choice in the UK Audit Market*, London: Oxera Consulting.

Peel, M.J. and Roberts, R. (2003). 'Audit fee determinants and auditor premiums: evidence from the micro-firm submarket'. *Accounting and Business Research*, 33(3): 207–233.

Pittman, J.A. and Fortin, S. (2004). Auditor choice and the cost of debt capital for newly public firms'. *Journal of Accounting and Economics*, 37: 113–116.

Pong, C. and Whittington, G. (1994). 'The determinants of audit fees: some empirical models'. *Journal of Business Finance & Accounting*, 21(8): 1071–1095.

Puhani, P.A. (2000). 'The Heckman correction for sample selection and its critique'. *Journal of Economic Surveys*, 14: 53–68.

Rosenbaum, P. and Rubin, D. (1983). 'The central role of the propensity score in observational studies for causal effect'. *Biometrika*, 70(1): 41–55.

Seetharaman, A., Gul, F.A. and Lynn, S.G. (2002). 'Litigation risk and audit fees: evidence from UK firms cross-listed on US markets'. *Journal of Accounting and Economics*, 33: 91–115.

Simonsen, M. and Skipper, L. (2006). 'The costs of motherhood: an analysis using matching estimators'. *Journal of Applied Econometrics*, 21(7): 919–934.

Simunic, D. (1980). 'The pricing of audit services: theory and evidence'. *Journal of Accounting Research*, 18(1): 161–190.

Simunic, D. and Stein, M.T. (1996). 'The impact of litigation risk on audit pricing: a review of the economics and the evidence'. *Auditing: A Journal of Practice and Theory*, 15: 119–134.

Stock J.H. and Watson, M.W. (2003). *Introduction to Econometrics*. Boston, MA: Addison Wesley.

Stock J.H., Wright, J.H. and Yogo, M. (2002). 'A survey of weak instruments and weak identification in generalised method of moments'. *Journal of Business and Economic Statistics*, 20(4): 518–529.

Stolzenberg, R.M. and Relles, D.A. (1997). 'Tools for intuition about sample selection bias and its correction'. *American Sociological Review*, 62: 494–507.

Titman, S. and Trueman, B. (1986). 'Information quality and the valuation of new issues'. *Journal of Accounting and Economics*, 8: 159–172.

Tucker, J. (2007). 'Is openness penalized? Stock returns around earnings warnings'. *Accounting Review*, 82(4): 1055–1087.

van der Klaauw, B. and Koning, R.H. (2003). 'Testing the normality assumption in the sample selection model with an application to travel demand'. *Journal of Business and Economic Statistics*, 21(1): 31–42.

White, H. (1980). 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica*, (May): 817–838.

Winship, C. and Mare, R.D. (1992). 'Models for sample selection bias'. *Annual Review of Sociology*, 18: 237–250.