IRANIAN JOURNAL OF ACCOUNTING, AUDITING & FINANCE

Journal homepage: https://ijaaf.um.ac.ir
DOI: 10.22067/ijaaf.2020.39426
Research Article

Difference-in-differences Design and Propensity Score Matching in Top Accounting Research: A Short Guide for Ph.D. Students in Iran

Reza Hesarzadeh*

Faculty of Economics and Business Administration, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract

In recent years, to increase the robustness of methodology sections of accounting research, applying quasi-experimental methods has become a popular approach in archival-empirical research of top-tier accounting journals. The purpose of this study is to discuss the usefulness of the two most robust methods, including difference-in-differences (DD) and propensity score matching (PSM). This paper discusses DD and PSM design and reviews DD and PSM's use in articles of American Accounting Associations' journals in recent years. In addition to a simple explanation of DD and PSM, this research provides a list of credible empirical accounting studies that have used these two methods. The research also explores the reasons for using the two methods in the empirical-archival studies of accounting and shows that in addition to extracting a causal relationship, the most important reason for using the two methods is to reduce the potential concerns surrounding the "omitted variables" "and "heterogeneity of treatment and control groups". Overall, by highlighting the importance and application of the DD and the PSM, this research can help the methodology sections' robustness in the empirical-archive accounting research that focuses on causal relationships and provide a simple and practical guide, especially for Ph.D. students in accounting.[1]

Keywords: Quasi-experimental methods, Causal relationships, Difference-in-differences, Propensity score matching, Archival-empirical accounting research.

* ***Corresponding Author***: Associate Professor of Accounting, Email: hesarzadeh@um.ac.ir

1- This paper is presented in 17th National Accounting Conference in Qom.

1. Introduction

Concluding causal relationships is often the primary objective of archival-empirical accounting research [Gow et al., 2016]. For this reason, in recent years, applying quasi-experimental methods has become a popular approach in archival-empirical accounting research (for example, see Dutillieux et al., 2016; Gunn and Michas, 2017; Kraft et al., 2018). Among the quasi-experimental methods, the two robust methods, including difference-in-differences (DD) and propensity score matching (PSM), in recent years, has attracted a lot of attention in accounting research. The purpose of this paper is to discuss the usefulness of these two methodologies, especially for Ph.D. students who tend to focus on the causal relationships in their dissertations.

As previously mentioned, DD and PSM methods have become increasingly popular ways to estimate causal relationships. DD consists of identifying a specific intervention or treatment (often the passage of the law). One then compares the difference in outcomes after and before the intervention for groups affected by the intervention to the same difference for unaffected groups. For example, to identify the incentive effects of specific disclosure regulation, one might first isolate firms under that regulation. Then compare changes in a dependent variable such as earnings management, for firms are under that regulation to the firms are not under that regulation. The great appeal of DD comes from its simplicity and its potential to circumvent many of the endogeneity problems that typically arise when making comparisons between heterogeneous individuals [Meyer, 1995]. DD has been widely used when evaluating a given intervention entails collecting panel data or repeated cross-sections. DD integrates the fixed effects estimators' advances with the causal inference analysis when unobserved events or characteristics confound the interpretations [Angrist and Pischke, 2009]. Whether serial correlation has led to a severe overestimation of t-statistics and significance levels in the DD literature so far depends on (1) the typical length of the time series used and (2) the serial correlation of the most commonly used dependent variables [Conley and Taber, 2011]. Further, DD is relevant for various cases where spillovers may occur between quasi-treatment and quasi-control areas in a (natural) experiment.

PSM is a matching technique that attempts to estimate the effect of a policy or other intervention by accounting for the covariates that predict receiving the treatment. PSM is for cases of causal inference and sample selection bias in empirical settings in which few units in the non-treatment comparison group are comparable to the treatment units or selecting a subset of comparison units similar to the treatment unit is difficult because units must be compared across a high-dimensional set of pre-treatment characteristics (Imai et al., 2004). **PSM** creates sets of participants for treatment and control groups. A matched set consists of at least one participant in the treatment group and one in the control group with similar propensity scores. The goal is to approximate a random experiment, eliminating many of the problems with observational data analysis.

Overall, in addition to a simple explanation of DD and PSM's method, this research provides a list of credible empirical accounting studies that have used these two methods. The research also explores the reasons for using the two methods in the empirical-archival studies of accounting and shows that in addition to extracting a causal relationship, the most important reason for using the two methods is to reduce the potential concerns surrounding the omitted variables and heterogeneity of treatment and control groups.

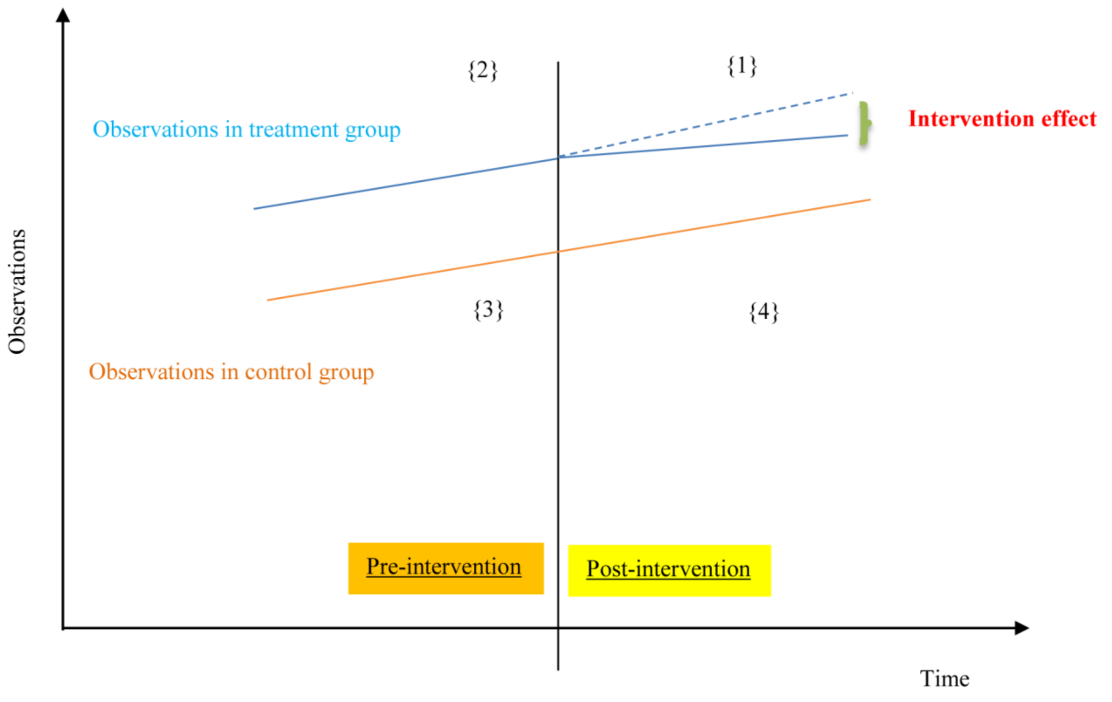
The paper's remainder is organized as follows: Section 2 discusses the DD methodology and Section 3 discusses the PSM methodology. Section 4 summarizes the study.

2. DD Method

Academic accounting researchers are often interested in interventions such as new policies like e new accounting standards and thought event studies. Simple event studies usually suffer from many variables that cannot be captured. Thus, a simple solution for mitigating this concern is randomization. In capital market settings, randomization to firms is unfeasible, and researchers are left with the need to use non-experimental studies to estimate the effects of these interventions. The fundamental challenge in such non-experimental studies is selection bias, in the sense that the firms experiencing the policy of interest may be different from those not exposed to it (Dutillieux et al., 2016). For example, firms that choose to apply a new standard may be quite different (and serve patients quite different) from those that do not apply. A common non-experimental design used to estimate the effects of policies at a particular point in time is a DD. DD compares changes over time in a group unaffected by the policy change to changes in a group affected by the policy change and attributes the differences to the policy's effect. DD provides unbiased effect estimates if the trend over time would have been the same between the treatment (intervention) and comparison groups in the intervention's absence. Because of information on the comparison group's temporal trends, DD is sometimes preferred over interrupted time series designs that do not necessarily have a comparison group.

Regarding the DD background, the first study using explicitly a DD is the (Snow, 1855). Snow (1855) was interested in the question of whether cholera was transmitted by (bad) air or (bad) water. He used a change in the water supply in one district of London, i.e., the switch from polluted water taken from the Themes in London's center to a supply of cleaner water taken upriver. Later on, the DD became relevant for other fields, like economics. For example, [Obenauer and von der Nienburg, 1915] analyzed the effect of a minimum wage by introducing the minimum wage for a particular group of employees, which led to higher wage rates in Portland, the largest city, compared to the rest of the state. Therefore, they documented the levels of various outcome variables for the different groups of employees in Portland before and after introducing the minimum wage and compared the respective changes to those computed for Salem, located in Oregon and thought to be comparable to Portland. Over time the field of economics developed literature. DD has been used to address many other important policy issues, like the effects of minimum wages on employment (e.g., Card and Krueger, 1994), or the effects of training and other active labor market programs for unemployed on labor market outcomes (e.g., Blundell et al., 2004).

DD may be a good choice when using research designs based on controlling for confounding variables or using instrumental variables is deemed unsuitable. At the same time, pre-treatment information is available. In many applications, “time” is an important variable to distinguish the groups. Figure 1 illustrates the DD. Besides the group which already received the treatment (post-treatment treated) {1}, these groups are the treated prior to their treatment (pre-treatment treated) {2}, the nontreated in the period before the treatment occurs to the treated (pre-treatment nontreated) {3}, and the nontreated in the current period (post-treatment nontreated) {4}.



Particularly, DD is used in settings where exchangeability cannot be assumed between the treatment and control groups; i.e., in the absence of treatment, the unobserved differences between treatment and control groups are the same over time. Hence, DD is a useful technique to use when randomization on the individual level is not possible. DD requires data from pre-/post-intervention, such as panel data (individual-level data over time) or repeated cross-sectional data (individual or group level). The approach removes biases in post-intervention period comparisons between the treatment and control groups that could result from permanent differences between those groups and biases from comparisons over time in the treatment group that could result from trends due to other causes.

Although other plausible methods are based on the availability of observational data for causal inference, i.e., instrumental variable, DD offers an alternative to reaching the un-confoundedness by controlling for unobserved characteristics and combining it with observed or complementary information. Additionally, the DD is a flexible form of causal inference because it can be combined with other procedures, such as the Kernel Propensity Score (Heckman, 1998).

Technically, to capture the effects in Figure 1, the regression below should be generated:

Dependent $\text{Variable}_{i,t} = \gamma_0 + \gamma_1$ Treatment-$\text{ControlGroup}_{i,t} + \gamma_2 \text{Post}_{i,t} + \gamma_3$(Treatment-$\text{ControlGroup}_{i,t} \times \text{Post}_{i,t}) + \sum \varphi$ (Controls)

Where *Treatment-ControlGroup* is set equal to one for the treatment group and zero for the control group. The coefficient of interest is γ_3, representing the differential change in the *Dependent Variable* between the treatment group and the control group. Controls are the control variables obtained from theory or prior studies.

An important assumption of the DD methodology is that shocks contemporaneous with the comment letters affect the treatment and control groups similarly (Johnston and Petacchi, 2017). To examine this assumption, a common way is to compare important variables for the treatment group and the matched control group. In the next section, I discuss more strong ways to examine the assumption.

DD has become a popular technique for concluding causal relationships in accounting research. Figures 2 and 3 present the relevant recent studies in the American

Accounting Associations' journals from 2016–2018. Specifically, Figure 2 overviews the studies that use DD, and Figure 3 overviews the reasons which explain why the studies use DD. Briefly, I find 17 studies that use DD from 2016 to 2018. Furthermore, the reason which the studies most refer is mitigating the concerns over omitted variables. For example, Kraft et al. (2018) discuss that the staggered timing of the change in reporting frequency gives us a natural group of control firms to implement a DD design in which they compare the change in investments of treatment firms around a reporting frequency increase relative to the contemporary change in investments for the control firms with unchanged reporting frequency. Therefore, they conclude that DD mitigates concerns about the effect of unobserved common shocks or cross-sectional differences across firms. Besides, Dutillieux et al. (2016) argue that the advantage of the DD design is that each sample firm acts as its own control over the test period, mitigating the concern for omitted correlated variables.

3. PSM Method

PSM is a statistical matching technique that attempts to estimate a treatment's effect by accounting for the covariates that predict receiving the treatment. PSM is for cases of causal inference and sample selection bias in non-experimental settings in which: few units in the non-treatment comparison group are comparable to the treatment units, or selecting a subset of comparison units similar to the treatment unit is difficult because units must be compared across a high-dimensional set of pre-treatment characteristics (Imai and Van Dyk, 2004).

PSM creates sets of participants for treatment and control groups. A matched set consists of at least one participant in the treatment group and one in the control group with similar propensity scores [Lunceford and Davidian, 2004]. The goal is to approximate a random experiment, eliminating many of the problems with observational data analysis.

The possibility of bias arises because the apparent difference in outcome between these two groups of the sample may depend on characteristics that affected whether or not a sample received a given treatment instead of due to the effect of the treatment per se. In randomized experiments, the randomization enables unbiased estimation of treatment effects; for each covariate, randomization implies that treatment-groups will be balanced on average by the law of large numbers. Unfortunately, for observational studies, the assignment of treatments to research subjects is typically not random. It is matching attempts to mimic randomization by creating a sample of units that received comparable treatment on all observed covariates to a sample of units that did not receive the treatment (Shaikh et al., 2009).

For example, one may be interested to know the consequences of smoking or the consequences of going to university. The people 'treated' are simply those—the smokers or the university graduates—who, in everyday life, undergo whatever it is the researcher is studying that. In both cases, it is unfeasible (and perhaps unethical) to randomly assign people to smoke or university education, so observational studies are required. The treatment effect estimated by simply comparing a particular outcome—a rate of cancer or lifetime earnings—between those who smoked and did not smoke or attended university and did not attend university would be biased by any factors that predict smoking or university attendance, respectively (Shipman et al., 2016). PSM attempts to control for these differences to make the groups receiving treatment and not-treatment more comparable.

Figure 2. Recent studies in AAA's journals who use DD

Authors (Year)	Title	Journal
Anantharaman et al. (2016)	State Liability Regimes within the United States and Auditor Reporting	The Accounting Review
Cheng et al. (2016)	Internal Governance and Real Earnings Management	The Accounting Review
Dutillieux et al. (2016)	The Spillover of SOX on Earnings Quality in Non-U.S Jurisdictions	Accounting Horizons
Lennox (2016)	Did the PCAOB's Restrictions on Auditors' Tax Services Improve Audit Quality?	The Accounting Review
Li and Yang (2016)	Mandatory Financial Reporting and Voluntary Disclosure: The Effect of Mandatory IFRS Adoption on Management Forecast	The Accounting Review
Chen et al. (2017)	XBRL Adoption and Bank Loan Contracting: Early Evidence	Journal of Information Systems
Francis et al. (2017)	Auditor Changes and the Cost of Bank Debt	The Accounting Review
Honaker and Sharma (2017)	Does Schedule UTP Have Uniform Long-Run Effects on Corporate Tax Planning?	The Journal of the American Taxation Association
Huang et al. (2017)	Product Market Competition and Managerial Disclosure of Earnings Forecasts: Evidence from Import Tariff Rate Reductions.	The Accounting Review
Kim and Klein (2017)	Did the 1999 NYSE and NASDAQ Listing Standard Changes on Audit Committee Composition Benefit Investors?	The Accounting Review
Dong and Zhao (2018)	Do Firms Do What They Say? The Effect of the American Jobs Creation Act of 2004 on R&D Spending	The Journal of the American Taxation Association
Jiang et al. (2018)	Big N auditors and audit quality: New evidence from quasi-experiments	The Accounting Review
Amin et al. (2018)	The Effect of the SEC's XBRL Mandate on Audit Report Lags	Accounting Horizons
Kraft et al. (2018)	Frequent Financial Reporting and Managerial Myopia	The Accounting Review
Li et al. (2018)	The determinants and consequences of tax audits: Some evidence from China.	The Journal of the American Taxation Association
Zhou and Chen (2018)	XBRL Adoption and Systematic Information Acquisition via EDGAR	Journal of Information Systems

Figure 3. The reasons for recent studies in AAA's journals for using DD

Authors (Year)	**Why DD?**
Lennox et al. (2018)	I use a difference-in-differences design to exploit a quasi-exogenous regulatory shock to the auditor-provided tax services (APTS) banned by the PCAOB. In contrast, prior studies undertake cross-sectional comparisons of companies that spend relatively more (less) on APTS.
Li and Yang (2016)	No certain/specified explanation
Anantharaman et al. (2016)	To more cleanly gauge the causal role of this legislation, we rely on a difference-in-differences approach that compares changes in outcomes in NJ to those in New York (NY) over the same period since these two states are geographically close and economically similar.
Cheng et al. (2016)	… DD research design to address endogeneity ….
Dutillieux et al. (2016)	The advantage of the difference-in-differences design is that each sample firm acts as its control over the test period, mitigating the concern for omitted correlated variables.
Kim and Klein (2017)	Event studies, however, have several empirical drawbacks, including the assumption of an efficient semistrong market during our sample period. Or, the market, although efficient, may underestimate the net benefits of the new listing standard. We also estimate difference-in-differences regressions to assess changes in financial reporting quality surrounding the 1999 rule change phase-in period to assuage these concerns.
Huang et al. (2017)	Our difference-in-differences design further mitigates concerns that other concurrent events confound our results.
Honaker and Sharma (2017)	The primary advantage of using a difference-in-differences technique is that it tests potential changes in the dependent variable over time. Given an event, it is a statistically powerful alternative to a change specification analysis. Its specification allows a firm to serve as its own control. Any random changes in firm characteristics over time are controlled, which also controls for non-independence in the variables of interest. A difference-in-differences estimation also controls for unobserved heterogeneity between the pre-and post-Schedule UTP periods that are constant over time.
Francis et al. (2017)	Employing a difference-indifferences research design to address potential endogeneity ….
Dong and Zhao (2018)	The standard diff-in-diff model is a tool to estimate treatment effects comparing the pre-and post-treatment differences in the outcome of a treatment and a control group. Therefore, it is commonly used for policy evaluation.
Kraft et al. (2018)	The staggered timing of the change in reporting frequency gives us a natural group of control firms to implement a difference-in-differences (DD) design in which we compare the change in investments of treatment firms around a reporting frequency increase relative to the contemporary change in investments for the control firms with unchanged reporting frequency. This design mitigates concerns about the effect of unobserved common shocks or cross-sectional differences across firms.
Zhou and Chen (2018)	No certain/specified explanation
Jiang et al. (2018)	Because the Big N acquisitions occurred at different points in time, not at the same time, we use a staggered difference-in-differences (DID) research design to estimate whether there is a Big N effect. This approach is consistent with studies focusing on settings with staggered treatment events

Studies using non-experimental data must mitigate endogeneity concerns introduced by non-random treatment assignment. In this regard, archival studies use multiple regression models to mitigate endogeneity concerns in observational data. However, multiple regression requires proper specification of the relation between outcome and explanatory variables to obtain unbiased estimates. If the relation between outcome and explanatory variables is misspecified, multiple regression can produce biased estimates. This potential bias increases as treatment groups become more dissimilar (Garrido, 2014). The PSM alleviates these concerns by decreasing reliance on the specification of the relationship between variables.

Regarding the general process of PSM, there are main four steps to apply the PSM efficiently: (1) Run logistic regression, where Dependent variable: Y = 1, if participate or for example, Y is higher than the median; Y = 0, otherwise; and independent variables are variables hypothesized to be associated with both treatment and outcome. (2) Obtain propensity score by extracting the predicted value from the regression in the previous step. (3) Match each participant to nonparticipants by propensity score. (4) Verify that covariates are balanced across treatment and matched control groups of a sample. For example, Eshleman and Guo (2014) use a logit regression for estimating propensity scores. After obtaining the fitted values from the logit regression, they match each non-Big 4 clients to the Big 4 client with the closest fitted value in the same year and same two-digit SIC code industry, requiring a maximum distance of 0.01 between the two fitted values. Then, they provide a test of covariate balance between matched pairs.

Similar to DD, but somewhat fewer, PSM has become a popular technique for concluding causal relationships in accounting research. Figures 4 and 5 present the relevant recent studies in the American Accounting Associations' journals from 2016–2018. Specifically, Figure 4 overviews the studies that use PSM, and Figure 5 overviews the reasons which explain why the studies use PSM. Briefly, I find 12 studies that use PSM from 2016 to 2018. Furthermore, the studies most refer to mitigating self-selection bias concerns and increasing treatment and control groups' comparability. For example, Gunn and Michas (2017) discuss that First, about potential selection bias, clients who choose to be audited by an auditor with multinational and/or country-specific expertise may exhibit firm-specific characteristics correlated with both this choice and our outcome variable. We perform a propensity score matching procedure, which can help alleviate this concern to the extent that clients and auditors are matching observable. In addition, Kraft et al. (2018) state that they use propensity score matching to identify control firms' sets.

4. Conclusions

The DD and PSM designs for empirical analysis of causal effects have a long history in outside accounting. Nowadays, they are certainly the most heavily used empirical research designs to estimate the effects of policy changes or interventions in empirical business. It has the advantage that the basic idea is intuitive and easy to understand for an audience with limited education. Compared to other methods, they have a further advantage that there is no need to control all confounding variables. This means that it can accommodate a certain degree of selectivity based on unobservables correlated with treatment and outcome variables. Its key identifying assumption is the common trend assumption that must hold unconditionally or conditionally on some observables (the treatment does not influence that). If the latter is the case, DD can be combined fruitfully with matching estimation techniques to flexibly accommodate such covariates.

Figure 4. Recent studies in AAA's journals who use PSM

Authors (Year)	Title	Journal
Bills et al. (2016)	Small Audit Firm Membership in Associations, Networks, and Alliances: Implications for Audit Quality and Audit Fees	The Accounting Review
Lennox (2016)	Did the PCAOB's Restrictions on Auditors' Tax Services Improve Audit Quality?	The Accounting Review
Li and Yang (2016)	Mandatory Financial Reporting and Voluntary Disclosure: The Effect of Mandatory IFRS Adoption on Management Forecast	The Accounting Review
Dutillieux et al. (2016)	The Spillover of SOX on Earnings Quality in Non-U.S Jurisdictions	Accounting Horizons
Garg et al. (2017)	Evaluating the Credibility of Voluntary Internal Controls Certification	Journal of International Accounting Research
Navissi et al. (2017)	Business Strategy, Over- (Under-) Investment, and Managerial Compensation	Journal of Management Accounting Research
Francis et al. (2017)	Auditor Changes and the Cost of Bank Debt.	The Accounting Review
Gunn and Michas (2017)	Auditor Multinational Expertise and Audit Quality.	The Accounting Review
Chi and Shanthikumar (2017)	Local Bias in Google Search and the Market Response around Earnings Announcements.	The Accounting Review
Huang et al. (2017)	Product Market Competition and Managerial Disclosure of Earnings Forecasts: Evidence from Import Tariff Rate Reductions.	The Accounting Review
Bauer et al. (2018)	Supplier Internal Control Quality and the Duration of Customer-Supplier Relationships	The Accounting Review
Kraft et al. (2018)	Frequent Financial Reporting and Managerial Myopia	The Accounting Review

Figure 5. The reasons of recent studies in AAA's journals for using PSM

Authors (Year)	**Why PSM?**
Bills et al. (2016)	No certain/specific explanation
Lennox (2016)	… propensity score matching helps to align the observable characteristics of the treatment and control groups.
Li and Yang (2016)	A concern with using all firms in non-IFRS countries as the control is their comparability with firms in IFRS countries in terms of disclosure incentives. We employ a propensity-score-matching (PSM) technique to pair firms in treatment and control groups based on observable characteristics to address this concern.
Dutillieux et al. (2016)	We used a matched propensity score methodology to derive the matched sample
Garg et al. (2017)	ICFR certification firms do not randomly decide to provide certification disclosure, which can result in self-selection bias. Bronson, Carcello, and Raghunandan (2006) examine firms' characteristics in issuing voluntary management reports on internal controls and suggest that some firms are more likely to make voluntary internal control disclosure than others. In order to rule out the probability that the results documented in the study are driven by differential firm characteristics rather than ICFR disclosures, we replicate our analyses using propensity score matching (PSM) of the control sample to control for self-selection.
Navissi et al. (2017)	We further mitigate the correlated omitted variable issue for the compensation hypotheses with propensity score matching procedure
Francis et al. (2017)	No certain/specific explanation
Gunn and Michas (2017)	First, about potential selection bias, clients who choose to be audited by an auditor with multinational and/or country-specific expertise may exhibit firm-specific characteristics correlated with this choice and our outcome variable. We perform a propensity score matching procedure, which can help alleviate this concern to the extent that clients and auditors are matching on observable variables
Chi and Shanthikumar (2017)	Visibility may affect the market response around earnings announcements in ways that linear models will not sufficiently control for. To address this issue, we use Propensity Score Matching (PSM) to form firm-year matched pairs that are most similar along with the set of firm characteristics included in equation (2) (the "covariates") but are most dissimilar in terms of their local bias in Google search (%Local)
Huang et al. (2017)	As a robustness test, we also perform a matched sample test based on the propensity score matching method.
Bauer et al. (2018)	we use a propensity score-matched (PSM) design to support that our primary findings do not suffer misspecification of the functional form (i.e., our ICW treatment firms are dissimilar to our non-ICW control firms)
Kraft et al. (2018)	We use propensity score matching to identify the set of control firms.

In conclusion, both DD and PSM are seen as strong non-experimental study design options for researchers, specifically Ph.D. students, who tend to find a causal effect. However, by combining them, we may make even more robust inferences, taking advantage of both important study design elements.

These methods also have their drawbacks. For example, most of the debate around the validity of a DD revolves around the possible endogeneity of the laws or interventions themselves. Sensitive to this concern, researchers have developed a set of informal techniques to gauge the extent of the endogeneity problem. Regarding DD and PSM's connection, it is worth stating that a concern with DD is that the intervention groups may differ in ways related to their trends over time, or their compositions may change over time. In this regard, PSM is commonly used to handle this confounding in other non-experimental studies.

References

Anantharaman. D. Pittman, J.A. and Wans, N. (2016). State Liability Regimes within the United States and Auditor Reporting. *The Accounting Review*, 91(6),1545-1575. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2372993

Angrist, J.D. and Pischke, J.S. (2009), *Mostly Harmless Econometrics*. ISBN-13: 978-0691120355

Amin. K. Eshleman, J.D. and Feng, C. (2018). The Effect of the SEC's XBRL Mandate on Audit Report Lags. *Accounting Horizons*, 32(1), 1-27. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2989894

Blundell, R. Meghir, C. Costa Dias, M. and Van Reenen, J. (2004). Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association* 2(4), 569–606. https://www.jstor.org/stable/40004874

Bills. K.L. Cunningham, L.M. and Myers, L.A. (2016). Small Audit Firm Membership in Associations, Networks, and Alliances: Implications for Audit Quality and Audit Fees. *The Accounting Review*, 91(3), 767-792. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2379678

Bauer A.M. Henderson, D.and Lynch, D.P. (2018). Supplier Internal Control Quality and the Duration of Customer-Supplier Relationships. *The Accounting Review*, 93(3), 59-82. Doi: 10.2308/accr-51889

Chi. S.S. and Shanthikumar, D.M. (2017). Local Bias in Google Search and the Market Response around Earnings Announcements. *The Accounting Review*, 92(4), 115-143. Doi: 10.2308/accr-51632

Conley, T. and Taber, C. (2011). Inference with "difference in differences with a small number of policy changes. *Review of Economics and Statistics*, 93, 113–125. https://www.nber.org/papers/t0312

Card, D. and Krueger, A.B. (1994). Minimum wages and employment:A case study of the fast-food industry in New Jersey and Pennsylvania. *The American Economic Review, 90(5)*, 772–793. https://www.jstor.org/stable/2677856

Cheng. Q. Lee, J. and Shevlin, T. (2016). Internal Governance and Real Earnings Management. *The Accounting Review*, 91(4), 1051-1085. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2666117

Chen. G. Kim, J.B. Lim, J. and Zhou, J. (2017). XBRL Adoption and Bank Loan Contracting: Early Evidence. *Journal of Information Systems In-Press*, 32(2), Doi: 10.2308/isys-51688

Dutillieux. W. Francis, J.R. and Willekens, M. (2016). The Spillover of SOX on Earnings Quality in Non-U.S. Jurisdictions. *Accounting Horizons*, 30(1), 23-39. Doi: 10.2308/acch-51241

Dong. Q. and Zhao, X. (2018). Do Firms Do What They Say? The Effect of the

American Jobs Creation Act of 2004 on R&D Spending. *The Journal of the American Taxation Association*, 40(1), 87-107. https://doi.org/10.2308/atax-51879
Eshleman, J.D. and Guo, P. (2014). Do Big 4 Auditors Provide Higher Audit Quality After Controlling for the Endogenous Choice of Auditor?. *Auditing: A Journal of Practice & Theory*, *33(4)*,197-219. Doi: 10.2308/ajpt-50792
Francis. B.B. Hunter, D.M. Robinson, D.M. Robinson, M.N. and Yuan, X. (2017). Auditor Changes and the Cost of Bank Debt. *The Accounting Review*,92(3),155-184. http://commons.aaahq.org/posts/4c6c42de48
Gunn. J.L. and Michas, P.N. (2017). Auditor Multinational Expertise and Audit Quality. *The Accounting Review*, 93(4), Doi: 10.2308/accr-51925
Gow, I. D. Larcker, D.F. and Reiss, P.C. (2016). Causal Inference in Accounting Research. *Journal of Accounting Research*, 54(2), 110-135. https://doi.org/10.1111/1475-679X.12116
Garg. M. Gul, F.A. and Wickramanayake, J. (2017). Evaluating the Credibility of Voluntary Internal Controls Certification. *Journal of International Accounting Research*, 16(3), 91-117. https://doi.org/10.2308/jiar-51856
Garrido, M. (2014). Methods for Constructing and Assessing Propensity Scores. *Health Services Research*, 49 (5), 1701–20. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4213057/
Heckman, J.J. (1998). *Comment on Eissa: Labor supply and the economic recovery act of 1981. In: M. Feldstein and J. Poterba (eds.): Empirical Foundations of Household Taxation.* 34(2), 5–38. https://doi.org/10.1111/1911-3846.12297
Honaker. K. and Sharma, D.S. (2017). Does Schedule UTP Have Uniform Long-Run Effects on Corporate Tax Planning?. *The Journal of the American Taxation Association*, 39(2), 63-79.
Huang, Y., Jennings, R. and Yu, Y. (2017). Product Market Competition and Managerial Disclosure of Earnings Forecasts: Evidence from Import Tariff Rate Reductions. *The Accounting Review*, 92(3), 185-207. https://doi.org/10.2308/accr-51558
Imai, K. and Van Dyk, D.A. (2004). Causal Inference with General Treatment Regimes: Generalizing the Propensity Score. *Journal of the American Statistical Association*, 99 (467), 854-866. https://doi.org/10.1198/016214504000001187
Johnston, R. and Petacchi, R. (2017). Regulatory oversight of financial reporting: Securities and Exchange Commission comment letters. *Contemporary Accounting Research*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1291345&rec=1&srcabs=1892286&pos=10
Jiang. J.X. Wang, I.Y. and Philip Wang, K. (2018). Big N auditors and audit quality: New evidence from quasi-experiments. *The Accounting Review*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2617802
Kim, S. and Klein, A. (2017). Did the 1999 NYSE and NASDAQ Listing Standard Changes on Audit Committee Composition Benefit Investors?. *The Accounting Review*, 92(6), 187-212. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2901187
Kraft. A.G.Vashishtha, V. and Venkatachalam, M. (2018). Frequent Financial Reporting and Managerial Myopia. *The Accounting Review*, 93(2), 249-275. Doi: 10.2139/ssrn.2456765
Lennox. C. S. (2016). Did the PCAOB's Restrictions on Auditors' Tax Services Improve Audit Quality?. *The Accounting Review*, 91(5), 1493-1512. http://commons.aaahq.org/posts/25aaa151b9
Li . X. and Yang, H.I. (2016). Mandatory Financial Reporting and Voluntary

Disclosure: The Effect of Mandatory IFRS Adoption on Management Forecasts. *The Accounting Review*, 91(3), 933-953. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2172014

Lennox, C.S. Li. W. Pittman, J. and Wang, Z. (2018). The determinants and consequences of tax audits: Some evidence from China. *The Journal of the American Taxation Association*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2585466

Lunceford, J.K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects. A comparative study. *Statistical Methods*, 23(19), 2937–2960. https://doi.org/10.1002/sim.1903

Meyer, B.D. (1995). Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics*, 13(2), 151–161. https://doi.org/10.2307/1392369

Navissi. F.V.S. Khedmati, M. Lim, E. and Evdokimov, E. (2017). Business Strategy, Over- (Under-) Investment, and Managerial Compensation. *Journal of Management Accounting Research*, 29(2), 63-86. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2806967

Obenauer, M. and Von der Nienburg, B. (1915). Effect of minimum wage determinations in oregon. *Bulletin of the U.S. Bureau of Labor Statistics*, 176, Washington, D.C.: U.S. Government Printing Office.

Shaikh, A. M. Simonsen, M. Vydacil, E.J. and Yildiz, N. (2009). A Specification Test for the Propensity Score Using Its Distribution Conditional on Participation. *Journal of Econometrics*, 151 (1), 33-46. https://doi.org/10.1016/j.jeconom.2009.01.014

Shipman, J.E. and Swanquist, Q.T. and Whited, R.L. (2016). Propensity Score Matching in Accounting Research. *The Accounting Review*, 92 (1), 213-244. Doi: 10.2308/accr-51449

Snow, J. (1855). *On the Mode of Communication of Cholera*. 2nd edition. London: John Churchill.

Zhou, J. and Chen, G. (2018). XBRL Adoption and Systematic Information Acquisition via EDGAR. *Journal of Information Systems*, 33(2), Doi: 10.2308/isys-52140