Int. Journal of Economics and Management 14 (1): 145-155 (2020)

IJEM
International Journal of Economics and Management
Journal homepage: http://www.ijem.upm.edu.my

Natural Experiment in Accounting and Finance Research

KAH-YUN WONG[a], CHEE-WOOI HOOY[a*] AND FAUZIAH MD. TAIB[a]

[a]*School of Management, Universiti Sains Malaysia, Malaysia*

ABSTRACT

Accounting and finance empirical studies always faces endogeneity issues and there have been many approaches being applied to overcome by scholars. In recent years, especially in the corporate governance related researches, scholars start to adopt a more convincing approach that is the natural experiments method. Natural experiments hugely rely on external shock and some name it as event study. Approaches being discussed here are Intend-to-Treat and Local Average Treatment Effect measures, Difference-in-Differences methods, and Propensity Score Matching method which are commonly used in the accounting and finance empirical studies.

JEL Classification: C10, C90, M40

Keywords: Accounting and Finance; Difference-in-Differences; Intend-to-Treat and Local Average Treatment Effect; Natural experiments; Propensity Score Matching

Article history:
Received: 8 November 2019
Accepted: 12 April 2020

[*] Corresponding author: Email: cwhooy@usm.my

INTRODUCTION

Endogeneity is always a problem in econometrics, and many accounting and finance empirical literatures are troubled by endogeneity, especially in investigating corporate governance matters (Roberts and Whited, 2012). Hence, many scholars address the concerns of endogeneity in finance and accounting literatures by presenting the issue firmly on table. For name a few, in accounting literatures, they are Chenhall and Moers (2007) and Larcker and Rusticus (2007). And in finance literatures, they are Roberts and Whited (2012) and Brown, Beekes and Verhoeven (2011). The current textbook solutions to endogeneity include two-stage least squares (2SLS) regressions, instrumental variables (IV), differenced generalised method of moments (GMM), and system GMM[1].

It is found that much of the empirical corporate governance literature lacks formal theory (Hermalin and Weisbach, 2003). Yet corporate governance studies, of which there are many, showing a relationship between corporate governance and firm value, have impacted policy, for example, policy relating to board composition and executive compensation included in the Sarbanes-Oxley Act (SOX) in the USA. Hence, for the vast number of corporate governance studies showing a relation between governance and performance, even if rigorously designed, it is safe to say they are fraught with endogeneity problems, that is, the optimum level of corporate governance is endogenously driven at the firm level (Demsetz, 1983). Due to that, applying a more persuasive and event study like natural experiments will be able to eliminate such endogeneity problems.

This study analyses the different approaches of natural experiments that are suitable to be applied on accounting and finance research based on different needs and scenarios. Among the total of 13,461 papers from social sciences journals (includes accounting and finance) (from 2001 to 2011) that Atanasov and Black (2016) have surveyed, they found only 40 papers that studied specific shocks and utilisize natural experiments. This showed that the methodology has yet to be widely used but also has brought to our attention the uniqueness of it. Leatherdale (2019) highlighted three main essence of nature experiments – enhance the ability to assess policy by government, boost the intensity of testimony and data about a policy's efficacy and deliver prompt information. Hence, this study allows fellow researchers to explore and obtain more concrete research results of the immediate impacts brought by sudden changes of policies or unexpected scenarios that happen in the market by adopting a more diverse yet effective research methodologies. Researchers in the field of accounting and finance will not constraint their research due to the limited time-based data availability by adopting natural experiments approaches.

The remainder of this paper is structured into four sections. Theoretical review on natural experiments are presented immediately after this section. Then followed by the discussions on the methodologies that can be applied on accounting and finance research. Next, empirical analysis and robustness tests are presented. The final section concludes this study.

THEORETICAL REVIEW ON NATURAL EXPERIMENTS

It has been highlighted that many of the existing studies, especially in finance and accounting research, faces the issue of endogeneity. Hence, developing better theory and more compelling evidence to fully describe the relationship between insider ownership and performance is accordingly part of the solution and this is where natural experiments can be valuable (Gippel et al., 2015).

What is natural experiment? A natural experiment can be one of two things: (1) a randomized controlled trial (RCT) set up by the researcher in a natural setting to induce controlled variance; or (2) a naturally occurring state (event) resulting from a social or political situation and thus not intentionally set up by the researcher. This second type of experiment is often referred to as a quasi-experiment (Meyer, 1995). Proponents of experimental research argue that such research designs avoid a criticism often levelled at econometric studies, that is, they are based on questionable economic theories.

In questions of finance, randomized experiments are difficult to implement. For example, it is neither feasible nor ethical to conduct randomized trials of bank failures, corporate governance changes, or tax

[1] Roberts and Whited (2012) discuss in great detail the issue of endogeneity and how it relates to the corporate governance literature.

changes. Thus, there are relatively few examples of randomized trials in a natural setting in the finance literature.

Debatably, the next level of evidence includes natural experiments in naturally occurring states or cases (Leigh, 2010). This type of experiment gives better opportunities for researchers in the field of finance and accounting. A *naturally occurring state* often comes about from a social or political situation (Dunning, 2007) such as a government policy change. Natural experiments are not ‘true’ experiments and are sometimes referred to as ‘quasi’ experiments (Meyer, 1995). It is due to the so-called *naturally occurring state* is not purposely set up by the researchers and hence assignment of the treatment grouping is not done randomly. Such experiments are more like observational studies where the researcher cannot manipulate the environment, although the researcher must choose the comparison or control group. In these types of research design, control groups and treatment groups may differ in systematic ways other than in regard to the treatment. The researchers therefore have to be attentive about ruling out those effects.

One of the pioneer natural experiment case using DiD is Snow (1855). Snow (1855) studied the effect of clean water on cholera death. The natural experiment was done when Lambeth Company moved its water source supplied to fresh water site while Southwark and Vauxhall Company also supplying to the same neighbourhood but remained in a polluted site. When cholera reappeared in 1853-54, the death rates on the group that water supply from Lambeth Company was minimalised while the group that the water supply from Southwark & Vauxhall Company was a lot higher.

Some examples of naturally occurring events of interest to finance and accounting researchers may include legislated corporate governance changes (e.g., SOX), or tax changes across jurisdictions that can justifiably be seen as exogenous sources of variation in the explanatory variables of interest. Meyer (1995) gives the example of the Vietnam era draft mechanism, which depended on date of birth as an exogenous event to study the effects of military service on earnings. Research utilizing natural experiments is growing in the field of economics (Dunning, 2007), although it is not very common in finance and accounting.

Finance and accounting researchers may see an obvious roadblock to designing research around a natural event, that is, how often will issue yield themselves to the kind of actual randomization that Heider and Ljungqvist (2014), for example, exploit? Researchers though, may not be aware of the number of situations or naturally occurring events that could be used in natural experimental research and thus many more natural experiments may be available than researchers realize. Generally, studies making use of natural experiments are motivated by existing evidence and thus address issues fundamental to a discipline. All address issues fundamental in the disciplines. Using such events as exogenous factors in a model helps researchers to provide a constructive link between the real world and econometric methodology (Angrist and Pischke, 2010).

The advantage of using natural experiments, when well considered, is that they are an exogenous event. Audiences shall be convinced of the matter by the researcher. In addition, researches utilising those events make a strong case of results on causal interpretation. Meyer (1995) argues that the main contribution of such research designs is in providing an understanding of the source of the causal relationship. The disadvantage is the time spent collecting data particularly considering there are often a lot of events. Another problem is validating the ‘as if’ random assignment of the comparison and control groups (Dunning, 2007).

Gippel et al. (2015) have suggested a guide steps to an ideal approach of research that attempting natural experiments. The steps are as follow:

1. Prepare a strong foundation for design/methodology of the research;
2. Identify the natural event(s);
3. Verify the natural event is plausibly exogenous;
4. Comparison of treatment group pre and post event;
5. Determine the control or comparison group;
6. Comparison of treatment and control group;
7. Control for standard variables raised in the literature;
8. Decide if the magnitude of the effects is economically meaningful; and
9. Reversal of the initial event.

Upon knowing the background and benefits of natural experiments, researchers need to evaluate and decide the best natural experiment designs or approaches that suits best to their research scenarios. At the early 20th century, Meyer (1995) proposes two basic designs – '*one group before and after design*' and '*before and after with an untreated comparison group*'. Meyer (1995) also showed that the second design can be extended further to multiple treated groups or multiple untreated comparison groups or multiple period based on the research scenario. The second design is more well-known as the Difference-in-Differences (DiD) approach. Then it can be seen that researcher like Arping and Sautner (2010) utilising the DiD approach, as well as Propensity Score Matching (PSM) as robustness test, in their natural experiment on the corporate governance reform in Netherlands. Atanasov and Black (2016) also found that 74 shock-based research papers that were published in the period of years 2001 to 2011 utilises different natural experiment approaches individually or by combining at least two approaches. In the recent "Impact Evaluation in Practice" and the accompanied Technical Companion by Gertler et al. (2016), they introduce six different natural experiment approaches to suit different scenarios. Three of the approaches – Instrumental Variables for the estimation of intend-to-treat (ITT) and local average treatment effect (LATE), Difference-in-Differences (DiD) and Matching concept, are discussed in further details in the following sections.

THE METHODOLOGY

Intend-to-Treat (ITT) Treatment and Local Average Treatment Effect (LATE) Treatment

When performing natural experiments, it is important to identify if there is any non-compliance occur in both the treatment and control groups. In practice, whenever there is a new rule or reformation in code of corporate governance being implemented, it can be either mandatory or voluntary basis. Hence, especially when a rule is voluntary basis, which also means there will be no penalty or sanction for the party that does not comply to it. This led to the existence of groups that comply as well as non-compliance group. It is important to identify if there is any non-compliance occur in both the treatment and control groups. Non-compliance is generally referring to some participants do not follow the rules (noncompliance in the treatment group) or some participants that are not exposed to the rules but voluntarily following the rules (noncompliance in the control group). In the presence of noncompliance, different parameters can be estimated. It is an important distinction between 'intent-to-treat' (ITT) estimates and 'local average treatment effect' (LATE) estimates to be performed to find out the actual effect from the implementation of the new rule.

The estimated impact α is called the "intention-to-treat" estimate (ITT) when the basic formula is applied to those firms to whom the quota has been enforced, regardless of whether they actually comply it. The ITT is important for those cases in which we are trying to determine the average impact of a quota on the sample targeted by the code. By contrast, the estimated impact α is called the "local average treatment effect" (LATE) when the basic impact evaluation formula is applied to those firms to whom the quota has been enforced and who have actually complied. The ITT and LATE estimates will be the same when there is full compliance, that is, when all firms to whom a quota has been enforced actually comply it.

Principally, the calculation obtained for ITT estimate (δ_{ITT}) will be smaller than the LATE (δ_{LATE}) estimates. This is due to the significance of the estimates of ITT is steered by the compliance entities that are qualified to the rule or shock-event and complying to it. Presuming that the new rule brings positive effects, if lesser entities comply and participate in the treated group, the average results over the entire treatment group shall be relatively smaller, and same goes to the ITT estimates. Hence, ITT estimates is actually the portion of the compliant entities multiplied by the LATE estimates. In the event of full compliance in the treatment group, there will be no difference between the ITT and LATE estimates. The relationship between both the ITT and LATE estimates can be summarised as $\delta_{ITT} = \text{Comply} * \delta_{LATE.}$

Hence, it is very crucial for researchers to understand their data and scenario background in order to find out the actual effect of each cases that being studied. Besides that, Atanasov and Black (2016) mentioned that researchers that use shock-based instrumental variables should also perform difference-in-differences (DiD) too. It is because LATE presumes that the instrument influences the outcome only through the instrumental variables (IV). However, DiD is the other way round instead. The coefficient from the DiD on the shock dummy estimates the total effect of the shock on the dependent variables (outcome). Hence, DiD is similar to the ITT estimates. DiD method are discussed further is the following section.

Difference-in-Differences (DiD) Method

Whenever there is a sudden or unexpected change or enforcement to a policy, it is best to perform a 'natural experiment' that allow researchers and analysists to identify the impact of that specific intervention or 'treatment'. One of the widely used methods by scholars is difference-in-differences (DiD) method. The main attraction of DiD estimation is due to its simplicity as well as its potential to avoid many of the endogeneity problems that typically emerge when making comparisons between heterogeneous individuals[2].

Hence, identification of treatment and control groups is also important. It is because the quality of the control group determines the quality of the evaluation made on the impact of the policy. However, for DiD to be valid, the control group must accurately represent the change in outcomes that would have been experienced by the treatment group in the absence of treatment. To apply DiD, it is necessary to measure outcomes in the group that receives the program (the treatment group) and the group that does not (the control group), both before and after the program.

DiD can be displayed in three (3) simple ways – in 'box', graphically and in a regression.[3] First, DiD in a 'box' is shown in Table 1. In Table 1, there are two time periods (before and after the program), as well as two groups of entities (those exposed to the program and those not exposed to the program). Time period taking the value $t = 0$ at baseline, and $t = 1$ at follow-up. Then, exposure to the program denoted by taking the value $P = 1$ for treatment group, and $P = 0$ for control group. Hence, utilizing DiD method, effects of a new rule or 'shock' event can be calculated using the Equation (1):

$$DiD = [(\bar{Y}_{t=1}|P = 1) - (\bar{Y}_{t=0}|P = 1)] - [(\bar{Y}_{t=1}|P = 0) - (\bar{Y}_{t=0}|P = 0)] \quad (1)$$

Table 1 Summary of Difference-in-Difference Approach in a table

	Group affected by the policy change (Treatment group) ($D = 1$)	Group not affected by the policy change (Control group) ($D = 0$)
After the program starts ($t = 1$)	$Y_{i,t=1}\|P_i = 1$ (Point B)	$Y_{i,t=1}\|P_i = 0$ (Point D)
Before the program starts ($t = 0$)	$Y_{i,t=0}\|P_i = 1$ (Point A)	$Y_{i,t=0}\|P_i = 0$ (Point C)
Before-after comparison	$(\bar{Y}_{t=1}\|P = 1) - (\bar{Y}_{t=0}\|P = 1)$	$(\bar{Y}_{t=1}\|P = 0) - (\bar{Y}_{t=0}\|P = 0)$

Source: Technical Companion of Gertler et al. (2016)

Besides that, DiD estimator is commonly presented in a regression model. After the basic DiD regression is formulated, it can then be expanded depending on each research cases and frameworks. The basic DiD regression formula is presented as Equation (2).

$$Y_{igt} = \beta_1 P_i + \beta_2 t + \delta P_i t + \alpha_g + \theta_t + \varepsilon_{igt} \quad (2)$$

where Y_{igt} for an individual i at time t in group g (treatment or control), P a dummy variable equal to one (1) for treatment group and zero (0) for control group, big t is a dummy variable equal to one (1) for post period and zero (0) for pre-period, β_1, β_2 and δ are the regression coefficients to be estimated, α_g is a time-invariant group-level fixed effect taking up differences between the treatment and control groups that are time-invariant, θ_t is the time-invariant fixed effect of constant effects related to each period and ε_{igt} is error term.

When in actual research practice, besides the outcome of interest, researchers also observe other characteristics for treatment and control groups in both time periods. Hence, these characteristics are observed for each unit in each group and time period (X_{igt}), and can be added to regression model and Equation (2) will be extended to become Equation (3) as follow:

$$Y_{igt} = \beta_1 P_i + \beta_2 t + \delta P_i t + \beta_3 X_{igt} + \alpha_g + \theta_t + \varepsilon_{igt} \quad (3)$$

[2] Refer to Meyer (1995) for more information and overview.
[3] Gertler, P. J.; Martinez, S., Premand, P., Rawlings, L. B. and Christel M. J. Vermeersch, 2010, Impact evaluation in practice: Ancillary material, The World Bank, Washington DC (www.worldbank.org/ieinpractice).
[a] - Illustration inspired by Figure 2 and Figure 4 of Lipsey, Farran, & Durkin. (2018).

Then, with reference to Table 1, the difference between the before-after comparisons in the treatment and control groups (namely the difference-in-differences) becomes as follow and presented in Equation (4):

$$(Y_{i11} - Y_{i00}) - (Y_{i01} - Y_{i00}) = \delta + (X_{i11} - X_{I10} - x_{I01} + X_{i00}) + (\varepsilon_{i11} - \varepsilon_{i10} - \varepsilon_{i01} + \varepsilon_{i00}) \quad (4)$$

It is important to note that for DiD to be valid, the control group must accurately represent the change in outcomes that would have been experienced by the treatment group in the absence of treatment. Besides that, it is also crucial that when one uses the DiD method, one must assume that, in the absence of the program, the outcome in the treatment group would have moved in tandem with the outcome in the control group. If outcome trends are different for the treatment and control groups, then the estimated treatment effect obtained by DiD methods would be invalid, or biased. That is because the trend for the control group is not a valid estimate of the counterfactual trend that would have prevailed for the treatment group in the absence of the program. For example, if in reality outcomes for the control group expand slower than outcomes for the treatment group in the absence of the program, using the trend for the control group as an estimate of the counterfactual of the trend for the treatment group leads to a biased estimate of the program's impact; more specifically, one would overestimate the impact of the program.

In addition, the DiD method compares trends between the treatment and control groups. The trend for an individual is the difference in outcome for that individual before and after the program. By subtracting the before outcome situation from the after situation, the effect of all of the characteristics that are unique to that individual and that do not change over time is cancelled out. Interestingly, the formula cancelling out (or controlling for) not only the effect of observed time-invariant characteristics, but also the effect of unobserved time-invariant characteristics.

The application of difference-in-differences method can be widely seen in previous literatures, especially the field of sociology and economics. To name a few, Card (1990) applies DiD in studying the effect of the Muriel Boatlift that caused increased of low-educated labour supply in Miami by comparing the individuals in Miami and individuals in other cities that not affected by the Muriel Boatlift; Eissa and Liebman (1996) investigate the impact of tax reform in the United States – Tax Reform Act of 1986 by comparing the change in the labor supply of single women with children with change for single women without children; Jin and Leslie (2003) study was inspired by the law set by Los Angeles County in 1998 that hygiene quality grade cards to be displayed in restaurant windows by investigating the effect of an increase in information related to the goods quality to consumers on the corporations' selections of product quality.

Besides that, in recent years, scholars that study corporate governance also applying DiD in their research. For example, Arping and Sautner (2010), Ahern and Dittmar (2012), Dale-Olsen et al. (2013), Bøhren and Staubo (2014), Fauver et al. (2017) and while Matsa and Miller (2013) attempting triple-differences.

EMPIRICAL ANALYSIS AND ROBUSTNESS TESTS

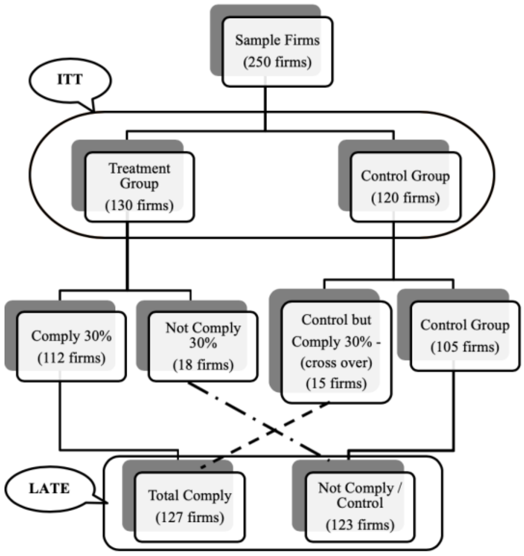
Sampling and Estimation Procedure of ITT Approach

Upon understanding the methodology of ITT and LATE estimates, one can put it into practice, especially for research analysis where non-compliance is foreseeable. A simple application and explanation of ITT estimates and LATE estimates is displayed in Figure 1. After inspecting and ensuring the entities fulfill the criteria of a newly implemented rules, 130 firms are identified as Treatment group, while 120 firms that are unaffected by the rules are categorized as Control group; this total to 250 firms selected for a study using natural experiment approach. These two groups are also identified as intent-to-treat (ITT) treatment and control groups for analysis, also known as ITT estimates. At this level, it is assumed that full compliance by the Treatment group and no entity from the Control group that follows the new rules. A simple regression framework is performed to obtain the ITT estimates effect on the new rules.

However, after detailed data collection, there are 18 firms in the treatment group that unable to achieve the required 30% rule (known as noncompliance in treatment group). While in the control group that not bound by 30% rule, there are 15 firms that have achieved the 30% rule, (known as crossover or noncompliance in control group). This resulted that from the total 250 firms, 127 firms are following the 30%

rule and the remaining 123 firms do not/no need to follow the 30% rule. These final two groups constitute the local average treatment effect (LATE) treatment and control groups respectively. Hence, another round of analysis, also known as LATE estimates, to be performed on LATE treatment and control groups to obtain the true effects of implementation of the 30% rule. Practically, instrumental variables approach is adopted to calculate the LATE estimates. After a valid instrumental variable is identified, the LATE estimates can be calculated using the 'two-stage least square' (2SLS) estimator.

In sum, it is important to first identify the treatment group and control group in a sample. Then, it is crucial to find out if there are any noncompliance matters. With that, higher accuracy of results from analysis of the effects of an implementation of a rules or regulation can be obtained. And to recap, ITT estimates the difference in outcomes between the units assigned to the treatment group and the units assigned to the comparison group, irrespective of whether the units assigned to the treatment group actually receive the treatment. While LATE estimates the difference in outcomes between the units that truly receive the treatment and the control group.



Note: Illustration adopted from Lipsey et al. (2018)[a] Total number of firms are for illustration purpose only to ease the ITT and LATE structures, assuming compliance rules in a case is minimum 30%.

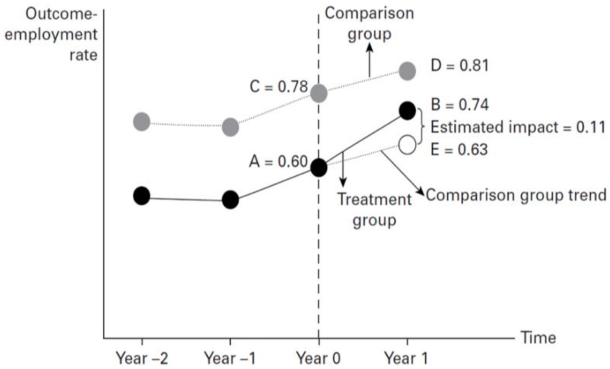
Figure 1 Illustration of Composition of Intent-to-Treat (ITT) and Local Average Treatment Effect (LATE) Groups

Results and Interpretation of the DID Approach

Next, it is important to be able to adopt and interpret the DiD approach empirically, which can be performed utilizing a graph as displayed in Figure 2. The graph can be visualized closely with Table 1 in the previous section. Points A to D seen in Figure 2 are reflecting the indication points A to D in Table 1.

In Figure 2, all differences between the points are to be interpreted as vertical differences in outcomes on the vertical axis. Year 0 is the baseline year, before the new program being implemented. In Year 1, treatment group enrols in the program, while comparison (control) group does not enrolled. The outcome level for the treatment group goes from point *A*, before the program starts, to point *B* after the program has started, while the outcome for the control group goes from point *C*, prior to the starting of the program, to point *D*, after the program has commenced. Using the change in outcomes for the control group as the estimate of the counterfactual for the change in outcomes for the treatment group is akin to assuming that, had the enrolled group not participated in the program, their outcome would have evolved over time along the same trend as the non-enrolled group: that is, the change in outcome for the enrolled group would have been from point *A* to point *E* (instead of to point *B*). Hence, the DiD impact shown in the graph can be computed as (Equation (5)):

$$DiD\ Impact, E = (B - A) - (D - C) \quad (5)$$



Source: Gertler et al. (2016)
Remarks: Figures in the graph for illustration purpose only.

Figure 2 The Difference-in-Differences Method

Once again, as a recap and as seen in Figure 2, it is a very important assumption that needed to be where before any new rules/program or code of corporate governance reformation, the trend for both the Treatment and Control groups are the same, moving in a parallel direction (refer to Year -2 to Year 0 in Figure 2). Hence, DiD approach comes in very useful whenever there are new things being implemented on the Treatment group. With this approach, we can see that the actual impact of program is not the movement of point *A* to point *B* (0.74 – 0.60 = 0.14) (usual scene when not applying natural experiment – DiD approach), but it is the actual 'unseen' impact between point *B* and point *E*. This actual impact after implementation of the new rules/program on Treatment group can be obtained after applying the Equation (5), that is amounted to 0.11 [(0.74 – 0.60) – (0.81 – 0.78)].

In sum, DiD approach is beneficial for finding out the impact of a new rules or a sudden event even with a short post-period data availability. DiD approach can be put into practice using multivariate regression framework. Besides that, as mentioned earlier, it is also worth noting that DiD does not take into consideration of the actual compliance or participation rate in the Treatment group. This is the reason that DiD is almost similar with the ITT estimates.

Robustness Checking: Propensity Score Matching (PSM) Method

As can be seen in many empirical studies, baseline model and DiD method would use cross-sectional data and to be conducted using panel regressions to achieve the research objectives and research hypotheses. However, it has always been criticized with selection bias issue, may be unreliable, and also does not necessarily represent the causal relationship between the compliance of the rule and the dependent variable. Therefore, as a robustness test, propensity score matching (PSM) (Rosenbaum and Rubin, 1983; Guo and Fraser, 2015) is used to solve the matters on selection bias by matching firms in the treatment group and control group but with otherwise similar observable characteristics.

In PSM, each unit in the treatment category and in the group of non-enrolled, to calculate the probability that this unit will participate or comply in the program (the so-called propensity score) based on the traits of the observed values (the explanatory variables). This score is a real number between 0 and 1 that summarizes the influence of all of the observed characteristics on the likelihood of complying in the event. Only baseline observed characteristics being used to compute the propensity score. This is because post-treatment characteristics could be influenced by the program itself and using such characteristics to verify the matched comparison group would bias the results. When the treatment affects individual characteristics and we use them to match, selection of a comparison group that seen akin to the treated group because of the treatment itself. Without the treatment, those characteristics would appear differently. This contravenes the basic requirement for a good estimate of the counterfactual, that is the control (comparison) group must be similar in all aspects, except for the fact that the treatment group receives the treatment and not the control group.

After the propensity score being computed for all units, units in the treatment group is used to match with items in the pool of non-enrolled that have the closest propensity score. These closest units used to form as the control group and are utilised to generate an estimate of the counterfactual. The propensity score–

matching method attempts to mimic the randomized assignment to treatment and control groups by selecting for the control group that have similar propensities to the entities in the treatment group.

Then, the average difference in results between the treatment or complied entities and their matched comparison entities delivers the estimated impact of the event or 'shock'. In sum, the event's effects and influences are estimated by comparing the average outcomes of a treatment group and the average outcomes among a statistically matched subgroup of units, the match being based on observed characteristics available in the data. In practice, it may be the circumstances where for certain complied entities, no entities in the pool of non-compliances have similar propensity scores. In technical terms, there may be a lack of common support, or lack of overlap, between the propensity scores of the treatment group and those of the pool of non-compliance.

There are three very important matters relating to matching. First, matching methods only observed characteristics to be used to construct a control group, due to the unobserved characteristics cannot be taken into consideration as it might affects the outcome and the estimation of effects computed with the matched control group would be biased. Next, matching must be performed using only characteristics that are not affected by the event/shock, for example variables such as gender, age. Third, the matching method's estimation results are only as good as the characteristics that are used for matching. It is crucial to do the matching on the basis of the attributes that determine compliance. The better understanding of the principal and benchmark of compliant selection, the easier it is to compute the matched control group.

As mentioned earlier, availability of baseline data on outcomes will enable matching to be combined with DiD approach to minimize the risk of bias in the computation. When propensity score matching (PSM) and DiD are merged in a calculation, any unobserved characteristics that constant across time between the treatment and control groups can be taken care of. It is carried out as follow:

i. Compute matching based on observed baseline traits,
ii. For every complied entity, compute the changes in outcome between the before and after events (first difference),
iii. For every complied entity, calculate the changes in outcome between the before and after events for the unit's matched control (comparison)(second difference),
iv. Use the DiD method - minus the second difference from the first difference,
v. Finally, average out all the double differences.

The above steps can be easily performed and obtained by using the 'psmatch2' command in Stata. Some of the scholars that utilizes PSM are Alkalbani and Mallin (2019); Doidge, Karolyi, and Stulz (2009); and Fauver et al. (2017).

CONCLUSION

After understanding the nature and benefits of natural experiments, scholars can adopt this approach by applying one or more of the natural experiments methods to their accounting and/or finance research to deliver a stronger and more convincing message to readers and policy makers. It is crucial to note that natural experiment approach reiterate that it is very crucial for the researcher to clearly comprehend the cause of variation utilized to estimate all the essential variables (Meyer, 1995). From guidelines by Gertler et al. (2016), there are a few different methods that can be utilized depending on the nature of the 'shock' event, data availability and compliances level.

In this review, we have covered three of the most frequently used natural experiments methods, especially DiD and PSM method. Both mentioned methods are commonly used in researching the topic of corporate governance, either purely on one country or few countries. However, the ITT and LATE method is less frequently seen, especially in the recent board gender diversity quota implemented in some of the Europe countries as the quota is mandatory (for example Norway) which means ITT and LATE will be the same results. But this method is especially useful on investigating rules or policies that are introduced but not mandatory.

In summary, natural experiments can be very useful in accounting and finance researches. This is because any sudden change in policies, revisions of tax rates, economic downturn or other events that happen from the external force permits a researcher to get a true exogeneous alternative in the main variables.

Utilizing this approach will truly enhance the value of the relationship found in any scenarios and gives more impact especially on event study despite short after-event period data availability.

ACKNOWLEGMENT

The authors would like to extend our appreciation to the Universiti Sains Malaysia for the Institute of Postgraduate (IPS) Grant (Grant No. 1001/PMGT/822120) that makes this paper publication possible.

REFERENCES

Ahern, KR & Dittmar, AK 2012, 'The changing of the boards: The impact on firm valuation of mandated female board representation', *The Quarterly Journal of Economics*, vol. 127, no. 1, pp. 137-197.

Angrist, J & Pischke, J 2010, 'The credibility revolution in empirical economics: How better research design is taking the con out of econometrics', *Journal of Economic Perspectives*, vol. 24, no. 2, pp. 3-30.

Arping, S & Sautner, Z 2010, 'Corporate governance and leverage: Evidence from a natural experiment', *Finance Research Letters*, vol. 7, no. 2, pp. 127-134.

Atanasov, V & Black, B 2016, 'Shock-based causal inference in corporate finance and accounting research', *Critical Finance Review*, vol. 5, pp. 207-304.

Bøhren, Ø & Staubo, S 2014, 'Does mandatory gender balance work? Changing organizational form to avoid board upheaval', *Journal of Corporate Finance*, vol. 28, pp. 152-168.

Brown, P Beekes, W & Verhoeven, P 2011, 'Corporate governance, accounting and finance: A review', *Accounting & Finance*, vol. 51, no. 1, pp. 96-172.

Card, D 1990, 'The impact of the Mariel boatlift on the Miami labor market', *ILR Review*, vol. 43, no. 2, pp. 245-257.

Chenhall, R & Moers, F 2007, 'The issue of endogeneity within theory-based, quantitative management accounting research', *European Accounting Review*, vol. 16, no. 1, pp. 173-96.

Dale-Olsen, H, Schøne, P & Verner, M 2013, 'Diversity among Norwegian boards of directors: Does a quota for women improve firm performance?', *Feminist Economics*, vol. 19, no. 4, pp. 110-135.

Demsetz, H 1983, 'Structure of ownership and the theory of the firm', *Journal of Law and Economics*, vol. 26, pp. 375-390.

Dunning, T 2008, 'Improving causal inference: Strengths and limitations of natural experiments', *Political Research Quarterly*, vol. 61, no. 2, pp. 282-293.

Eissa, N & Liebman, JB 1996, 'Labor supply response to the earned income tax credit', *The Quarterly Journal of Economics*, vol. 111, no. 2, pp. 605-637.

Fauver, L, Hung, M, Li, X & Taboada, AG 2017, 'Board reforms and firm value: Worldwide evidence', *Journal of Financial Economics*, vol. 125, no. 1, pp. 120-142.

Gertler, PJ, Martinez, S, Premand, P, Rawlings, LB & Vermeersch, CMJ 2016, *Impact evaluation in practice (2nd Ed)*. Washington, DC: Inter-American Development Bank and World Bank. © World Bank. https://openknowledge.worldbank.org/handle/10986/25030. License: Creative Commons Attribution CC BY 3.0 IGO.

Gippel, J, Smith, T & Zhu, Y 2015, 'Endogeneity in accounting and finance research: Natural experiments as a state-of-the-art solution', *Abacus*, vol. 51, pp. 143-168.

Guo, S & Fraser, MW 2015, *Propensity score analysis: Statistical methods and applications* (2nd ed.). SAGE Publications, Inc.

Heider, F & Ljungqvist, A 2014, As certain as debt and taxes: Estimating the tax sensitivity of leverage from state tax changes, Working Paper, New York University.

Hermalin, BE & Weisbach, MS 2003, 'Board of directors as an endogenously determined institution: A survey of the economic literature', *Federal Reserve Bank of New York Economic Policy Review*, vol. 9, pp. 7-26.

Jin, GZ & Leslie, P 2003, 'The effect of information on product quality: Evidence from restaurant hygiene grade cards', *The Quarterly Journal of Economics*, vol. 118, no. 2, pp. 409-451.

Larcker, DF & Rusticus, T 2007, 'Endogeneity and empirical accounting research', *European Accounting Review*, vol. 16, no. 1, pp. 207-215.

Leatherdale, ST 2019, 'Natural experiment methodology for research: a review of how different methods can support real-world research', *International Journal of Social Research Methodology*, vol. 22, no. 1, pp. 19-35.

Leigh, A 2010, 'Evidence-based policy: Summon the randomistas?', in P. Commission (Ed.), Strengthening Evidence Based Policy in the Australian Federation: Proceedings, Roundtable Proceedings, Productivity Commission, Canberra, pp. 215-226.

Matsa, D & Miller, A 2013, 'A female style in corporate leadership? Evidence from quotas', *American Economic Journal: Applied Economics*, vol. 5, no. 3, pp. 136-169.

Meyer, B 1995, 'Natural and Quasi-Experiments in Economics', *Journal of Business & Economic Statistics*, vol. 13, no. 2, pp. 151-161.

Roberts, M & Whited, T 2012, 'Endogeneity in empirical corporate finance', *Simon School Working Paper*.

Rosenbaum, PR & Rubin, DB 1983, 'The central role of the propensity score in observational studies for causal effects', *Biometrika*, vol. 70, no. 1, pp. 41-55.

Snow, J 1855, *On the mode of communication of cholera* (2nd ed.). London: John Churchill. Retrieved from: https://books.google.com.my/books/about/On_the_Mode_of_Communication_of_Cholera.html?id=-N0_AAAAcAAJ&redir_esc=y