The failure of the balanced condition in the natural experiment design

By Chen Wang*
Research School of Finance, Actuarial Studies and Applied Statistics, The Australian National University
*Corresponding student author. E-mail: chen.wang@anu.edu.au

ABSTRACT

One important task in economics and finance studies is establishing causal inference. Given the obstacles of obtaining reliable instrumental variables, methodologies such as ordinary least squares (OLS) that fall under the umbrella of natural experiments by taking advantage of exogenous events are becoming prominent, without questioning on the balanced condition hypothesis. One empirical problem with these methodologies is that the treatment assignment is not random, which is characterized by non-balanced covariates across the treatment and control groups. This problem is often not obvious to researchers, and they may infer causality when in fact none may exist. By employing the examples from influential journals, we show that the causal inferences from the natural experiment studies may change after we deal with the imbalanced condition problem. We argue that the data quality will affect the estimates of the treatment effect: a better-balanced dataset will require fewer matching activities, in addition to that, the estimation results after dealing with the imbalanced condition problem become less volatile compared with those from low quality dataset. Notwithstanding the popularity of nature experiment technique, according to our knowledge, such results are not available in the previous literature.

Introduction

Given the challenges of obtaining reliable instrumental variables to answer the causality of interest, researchers have increasingly exploited natural experiments. In a natural experiment, the assignment of treatment to subjects is serendipitous and random (Freedman, Pisani, and Purves, (1997)). Such alleged natural experiments are generally variations in rules controlling behaviour that are supposed to satisfy random criteria.

Unfortunately, as several commonly used methods (e.g., OLS) are not designed to focus on balance checking for covariates in studies with binary treatments, numerous applications have failed to determine the potential problem of distinct characteristics between treatment and control groups, which is the central dilemma—that is, the randomness assumption (or the balanced condition assumption) in a natural experiment does not hold. Such an imbalanced condition issue perhaps explains why "as if" "causality relationship" can be obtained by such a problematical "natural experiment" in which indeed such causal inference may be different from the original study since the groups compared are fundamentally heterogeneous in terms of their characteristics.

Proceeding without balance checking is a mistake as the criterion of randomness in the nature experiment is not satisfied. Thus, the researcher is left to worry that the "causality" established from the regression method (without a balance-checking process) may produce imprecise estimations, since it may actually stem from the heterogeneous characteristics in the samples compared.

We review 372 empirical natural experiment studies in Economics, Finance, and Accounting journals, only 152 (40.86%), 65 (17.52%) and 31 (8.36%) papers check the first moment, higher moments, and joint imbalanced problems respectively. In our reviewed 372 natural experiment studies, conditional on the Propensity Score Matching (PSM) method included in the papers, only 10.53% of the papers check the balanced condition required by the PSM method, in doing so, the potential estimation problem may be a concern as we illustrate in the Monte-Carlo section and the empirical study section.

We examine different natural experiment studies claiming causal inference in influential economic, financial and accounting journals, and compare the estimation performance between their original results (without matching step) and the results after dealing with the imbalanced condition problem with our combined method (direct matching combined with regression).

Balance checking problem in the literature

In order to evaluate the balance checking problem in literature, we searched the empirical natural experiment articles from the Web of Science database (from 2000 to the first half year of 2018) using the following terms or variations: terms focusing on "Natural" or "Quasi" or "Quasi Natural" near experiment; "exogenous" near "shock" or "event". We manually read each article that satisfied above settings and exclude the papers focusing on Regression Discontinuity Design (RDD) approach, Instrument Variable (IV) approach, the time series models' approach, the review papers, the discussion or comments papers on previous studies, and the pure theory papers without empirical studies so that the sample of articles satisfy the method settings we argue above.

Table A review of balanced condition problem in empirical natural experiment literature

Strategy	Yes	No
Checked First moment Balanced condition of Natural Experiment	152	220
(Percentage)	40.86%	59.14%
Checked Higher moments (second, third and fourth moments)	65	307
Balanced condition of Natural Experiment (Percentage)	17.47%	82.53%
Checked Global (Jointly) Balanced condition of Natural Experiment	31	341
(Percentage)	8.33%	91.67%
Include PSM methods	38	334
(Percentage)	10.22%	89.78%
If PSM is included, is PSM Balanced Condition checked?	4	34
(Percentage)	10.53%	89.47%

Method- Combining an direct matching (EB) method and structure models as well as Combing an indirect matching (PSM) method and structure models

In the natural experiment design study, we consider a sample including n1 observations in the treatment group and n2 observations in the control group. Each observation I is subject to a treatment $D_i \in \{0,1\}$, where $D_i = 1$ means that observation i is in the treatment group, and $D_i = 0$ means that unit i is coded in the control group. X represents a set of covariates. X_i^k is the i'th observation for covariate k. We employ the direct matching scheme -Entropy Balancing (EB) method (Hainmueller (2012)).To obtain the weights w_i for each observation in the control group such that the treatment and control groups are balanced with respect to the covariates, we minimize the entropy equation as follows

$$\mathrm{H}(w^i) = \sum_{\{i|D=0\}} w^i \log \frac{w_i}{q_i} \quad (1)$$

which is an entropy distance function from Kullback (1959), where

$$q_i = \frac{1}{n_2} \quad (2)$$

and "n_2" is the number of units in the control group. The loss function can also be chosen from Cressie Read divergence family (Read and Cressie (1988)). The reason why Kullback function is preferred is as its property is more stable, given the misspecification environment ((Imbens, Spady, and Johnson (1998)).

The entropy distance function in equation (1) is minimized subject to the following three constraints that ensure balance:

$$\sum_{\{i|D=0\}} w_i\, c_r^k(X_i^k) = m_r^k, r \in 1,2,\ldots \mathrm{R} \quad (3)$$

$$\sum_{\{i|D=0\}} w_i = 1 \quad (4)$$

$$w_i \geq 0\ \forall_i \text{ such that D=0} \quad (5)$$

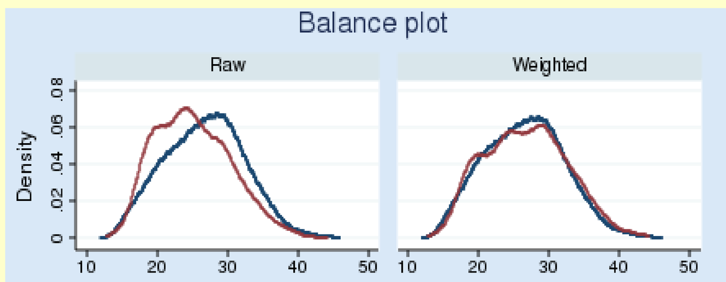
where the balance constraints $C_r^k(X_i^k) = m_r^k$ are imposed through a reweighting of the r moments for X_i^k of the control group. In the natural experiment studies, we include all the three moments (mean, variance, and skewness) as the constraints in the analysis if there is a solution on weights. Here, m_r^k is the moment's "r" for covariate k in the treatment group. C_r^k is the moment's function for covariate k in the control group. We can include all three moments' (mean, variance, and skewness) constraints, such that the mean, variance, and skewness in the control group are equal to those in the treatment group.

The typical estimation strategy in direct matching scheme is to calculate the mean differences between the treatment and control groups as the average treatment effect, without worrying about the true economic question where the economic structural models can provide the useful information of interest. Heckman and Urzua (2009) argue that even the perfect randomization experiment cannot answer the economic question of interest because of the lack of the economic mechanism analysis executed by structure models. In order to explore the benefits of both direct matching scheme and the economic structure models used in the natural experiment design, we try to explore ways to combine the direct matching method (EB) with the natural experiment method (a structure model is usually included within the natural experiment design) and to provide evidence that if this combined approach (called direct matching combined method thereafter) is appropriate in the non-randomness "natural experiment" study.

More specifically, we firstly produce the solution obtained from the equation (1) to equation (5), then the resulting weights is compatible with OLS, a commonly used approach in the natural experiment design.

Figure below reports an example of density function for a covariate with and without the weighting operation process. The left column reports the raw density function without a weighting operation. The right column reports the covariate density function after a weighting operation. The red curve represents the density function of the covariate in the control group. The blue curve represents the density function of the covariate in the treatment group. The figure shows that after the weighting operation, the covariates are rather similar between treatment and control groups.

Figure Density function with and without weighting operation process



Monte Carlo Simulation study

To illustrate the argument regarding what can go wrong if the covariates are not balanced in the compared group in the natural experiment's design, our investigation focuses on the Monte Carlo simulation where the covariates in the treatment and control groups are distinct so that the randomness assumption does not hold.

For intuition, we randomly draw two covariates for the control group (with 10,000 observations) from a uniform distribution (0,10) and two covariates for the treatment group (with 10,000 observations) from a uniform distribution (3,13); including more covariates does not change the implication and will make the argument even stronger. the overlapping observations as a [3,10]×[3,10] square as a random dataset, and the observations outside of this range provide imbalance to the dataset. In doing so, the covariates in the treatment group are fundamentally higher than those in control group, which is consistent with our assumption that the randomness condition does not hold in the "natural experiment". We then produce the dependent variable from the equation $z_i = y_{1i} + y_{2i} + \varepsilon_i$, where ε_i follows a standard normal distribution with mean of 0 and a standard deviation of 1. We repeat the simulation as suggested above for 100 times.

To evaluate the estimation performance produced by the direct matching (EB) scheme combined with OLS and the indirect matching (PSM) scheme combined with OLS, we use the mean difference from the [3,10]×[3,10] random experiment as a benchmark for the treatment effect, and we test which method can recover the benchmark treatment effect better if different model structure assumptions are used.

We first check the balanced condition for the unbalanced dataset and the result is reported in Table below. The table reports the imbalance problem between the treatment and control groups on the mean differences, the quantiles difference, and the L1 distance for joint covariates imbalance that is calculated by coarsening the data (Iacus, King and Porro (2009)).

Table: Balance checking for covariates between treatment and control groups

Balance Checking for covariates between treatment and control groups (Monte Caro Simulation)

Variable	Control Group						
	Treatment Group						
	$L1$	mean	min	0.25	0.5	0.75	max
y_{1i}	0.2833	3	3	3	3	3	3
y_{2i}	0.2827	3	3	3	3	3	3

Table: Estimates of the Monte Carlo Simulation Study

Compare direct matching (combined with OLS) with OLS (without matching scheme) and PSM (combined with OLS) in Monte Caro Simulation

Benchmark Settings	Benchmark Treatment Effect (Unadjusted)	Direct matching combined with OLS (Unadjusted)	Direct matching combined with OLS (Adjusted)	PSM combined with OLS (Unadjusted)	PSM combined with OLS (Adjusted)	OLS (without matching) (Unadjusted)	OLS (with matchin (Adjuste
Original Treatment Effects	0.044	0.036	0.035	0.019	-0.006	5.993	-0.008
Bias	0	0.008	0.009	0.025	0.050	5.949	0.052
No of observations	20,000	20,000	20,000	20,000	20,000	20,000	20,000

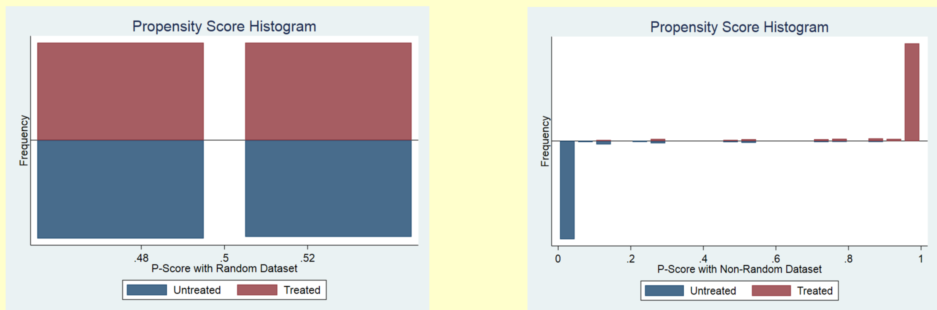
Compare the indirect matching PSM (combined with OLS) with OLS (without matching step) in Monte Carlo Simulation study

Possibly, the indirect matching scheme-PSM is the most popular strategy and commonly used in causal inference study (Pearl, 2010). The method has been used and referenced for around 116,000 research papers. We reanalyze the examples by using PSM method, an indirect matching scheme, because it is widely used in observational studies to deal with the problem of self-selection (non-randomness problem). According to Ho et al. (2007), Sekhon (2009), and Hainmueller (2012), PSM methods are commonly used for estimation of binary treatment effects based on the assumption of the selection of observables. One of the main criticisms of the PSM method is that, in practice, it can be challenging to ensure that the distributions of the propensity scores of the control and treatment groups are balanced (see for example Rosenbaum and Rubin, (1983, 1985); Dehejia and Wahba, (2002) for tests to check for balance). We investigate the balanced condition required by the PSM and extend the analysis in our studies to show that PSM may not produce both local balanced condition (if each individual covariate is the same between compared groups) and global balanced condition (if all the covariates are balanced jointly between treatment and control groups) if the analyzed data quality is low (the natural experiment is not random). We show here in the Monte Carlo Simulation study that PSM combined with OLS (called PSM combined method thereafter) may perform worse than OLS (without matching step) if the dataset suffers from serious imbalance problem in non-random dataset. In fact, it is difficult for PSM to achieve balanced condition if the dataset is too far from random, in doing so, PSM may not produce well balanced dataset to be analyzed.

We trace this problem in the Monte-Caro Simulation study. We create one random dataset, with two covariates for both control group and treatment group (both groups with 10,000 observations) from a uniform distribution (0,10). We also create one non-random dataset, with two covariates for the control group (with 10,000 observations) from a uniform distribution (0,10) and two covariates for the treatment group (with 10,000 observations) from a uniform distribution (8,18). All the other settings are following the same procedure as before.

We report the frequency distributions of the propensity scores for the two datasets in figure below. The left panel of the figure shows that PSM has achieved perfect balanced condition in random dataset, while the difficulty for the non-random dataset (right panel) is that most propensity scores in treatment group are much higher than that of control groups, in doing so, the balanced condition required by PSM is difficult to be met.

Figure : Frequency distributions of propensity scores



We show here the estimations from PSM combined method will be different from that estimated by OLS (without balancing checking step). The results, in the Panel A of table below show that PSM combined method initially does quite well regarding its estimation performance when the balanced condition is met in random dataset since its estimated parameter (1.981) is the same as OLS (1.981). However, as the balanced condition required by PSM deteriorates (as shown in bottom Panel of Figure 1), the PSM estimation performance (Panel B of table 5) suffers compared with OLS, and the potential MSE with PSM goes up dramatically compared with a small MSE of OLS (0.001)

Table: Compare PSM with OLS in the Monte Carlo Simulation Study.

Panel A: Random dataset results

Comparation between PSM method and OLS method (Random Dataset)

Benchmark Settings	OLS	PSM(Radius 0.01)	PSM(Radius 0.03)	PSM(Kernel Matching)
Treatment Effects	1.981	1.981	1.981	1.981
Observations	20,000	20,000	20,000	20,000

Panel B: Non-random dataset results

Comparation between PSM method and OLS method (Non-Random Dataset)

Benchmark Settings	OLS	PSM(Radius 0.01)	PSM(Radius 0.03)	PSM(Kernel Matching)
Treatment Effects	1.958	1.942	1.955	1.956
MSE	0.001	0.003	3.925	3.926
Observations	20,000	20,000	20,000	20,000

Application: The effects of Board Independence on CEO compensation

Recent empirical studies (Chhaochharia and Grinstein (2009); Duchin, Matsusaka, and Ozbas (2010); Chen (2014); Chen, Cheng, and Wang (2015); Armstrong, Core, and Guay (2014); Guthrie, Sokolowsky, and Wan (2012)) employ the "natural experiment" induced by the SEC regulation to study the effects of board independence.

Despite the strength of this design, that is, that firms which did not satisfy this requirement were serendipitously forced to increase their board independence level (thus, the variation in board structure of affected firms appears "exogenous"), this "natural experiment" lacks balance checking on the characteristics between compared groups in the study. Indeed, part of the predicament in this design is that the assignment of a treatment variable to subjects is not random, since we show that the firm characteristics are significantly different among compared groups.

Table: Balance checking for covariates in the study of board independence

Balance Checking for covariates between treatment and control groups

Variable	Control Group					Treatment Group					Difference
	Obs	Mean	Std.Dev.	Min	Max	Obs	Mean	Std.Dev.	Min	Max	
Sales*Dummy(00-02)	4,074	3.87	4.002	0	12.41	768	3.598	3.726	0	11.804	0.272*
Sales*Dummy(03-05)	4,074	3.936	4.063	0	12.636	768	3.715	3.84	0	11.892	0.221
ROA*Dummy (00-02)	4,074	0.024	0.054	-0.597	0.399	768	0.031	0.071	-1.209	0.283	-0.007***
ROA *Dummy (03-05)	4,074	0.019	0.083	-3.781	0.294	768	0.028	0.054	-0.314	0.316	-0.009***
Stock Returns*Dummy (00-02)	4,074	0.031	0.304	-1.852	2.089	768	0.043	0.324	-1.556	1.551	-0.012
Stock Returns *Dummy (03-05)	4,074	0.048	0.259	-2.198	1.318	768	0.055	0.257	-1.35	1.34	-0.007

We follow the designs presented in Chhaochharia and Grinstein (2009); the regression model is recorded as follows:

$$\mathrm{Log}(Compensation_{it}) = \alpha_0 + \alpha_1 \times Dummy(NoncompliantBoard02)_i \times Dummy(03-05)_t + \mathrm{Controls}_{it} + Firm\_Effects_i + Industry\_year\_Effects_{it} + \varepsilon_{it} \quad (6)$$

Table: Estimates of the board independence effect on CEO compensation

Estimates of the board independence effect on CEO compensation

Dependent Variable: Ln(Total CEO Compensation)	OLS (in the original natural experiment study)	Direct matching combined method
Independent Variables:	Coefficient	Coefficient
Dummy(Noncompliant Board(02)*Dummy(03-05))	-0.171*	-0.129
Sales*Dummy(00-02)	0.359***	0.144
Sales*Dummy(03-05)	0.325***	0.024
ROA*Dummy(00-02)	0.167	-0.035
ROA*Dummy(03-05)	0.21	0.963
Stock Returns*Dummy(00-02)	0.118***	0.063
Stock Returns*Dummy(03-05)	0.294***	0.172
CEO Tenure	-0.02	-0.064
Firm fixed effect	Yes	Yes
Industry-year fixed effect	Yes	Yes
Constant	Yes	Yes
Number of observations	4842	4842
Adjusted-R Square	0.643	0.697

Table: Comparison between the direct matching combined method and the OLS method (in the original natural experiment study)

Comparation between the direct matching combined method and the OLS method (in the original experiment study)

Benchmark Settings	Direct matching Combined method	OLS method (in the original experiment study)
Original Treatment Effects	-0.129	-0.171
Bias	0.031	0.048
MSE	0.019	0.221
No of simulations	1000	1000

Table: Comparison between the direct matching combined method and the PSM combined method

Comparation between the direct matching combined method and the PSM combined method

Benchmark Settings	Direct matching combined method	PSM combined method
Original Treatment Effects	-0.129	-0.193*
Bias	0.004	0.034
MSE	0.009	0.041
No of simulations	1956	5000

Table: Comparison between the direct matching combined method and various PSMs combined method

Comparison between the direct matching combined method and various PSMs combined method

Benchmark Settings	Direct matching combined method	Local Linear Regression Matching combined method	Radius Matching combined method (Radius=0.01)	Radius Matching combined method (Radius=0.02)	Radius Matching combined method (Radius=0.03)	Nearest Matching combined method
Original Treatment Effects	-0.129	-0.216	-0.180	-0.193	-0.197	-0.307
Bias	0.031	0.036	0.048	0.035	0.036	0.075
MSE	0.020	0.071	0.045	0.042	0.042	0.056
No of simulations	1000	1000	1000	1000	1000	1000

Table: Hotelling's T-Squared generalized means test and Rubin's B test

Hotelling's T-squared generalized means test and Rubins' B test

Statistics of the test	Hotelling's T-squared generalized means test (Null Hypothesis:Vectors of means are equal for the two groups)				Rubin's B
Raw Data	F(7,4834):	33.626	Prob>F (7,4834):	0	57.5*
Data After indirect matching	F(7,4821)	0.371	Prob > (7,4821)	0.920	4.6
Data After direct matching	F(7,4576):	0	Prob>F(7,4576):	1	0

Conclusion:

In this study, we provide recommendations to the researchers in the natural experiment area, and propose to apply the direct matching combined method when the data quality is low, but there is no distinct difference regarding causal inference among the direct matching combined method, the indirect matching combined method, and the natural experiment design (without matching step) in good quality dataset.