How Much Should We Trust Difference-in-Differences Estimates?—Empirical Evidence from Temporary Disability Insurance

Han Jiang

Abstract

Many empirical papers employ Difference-in Differences (DD) to estimate causal relationships by using many years of data. This model has a very important "parallel trend" assumption: the outcome in treatment and control group would follow the same time trend in the absence of the treatment. This paper evaluates the effects of paid maternity leave program in the 1978 Temporary Disability Insurance (TDI) on birth outcomes in the United States. Using NVSS (National Vital Statistics System) natality data, I find that DD does not perform well. Although all the results are statistically significant, it reveals that the TDI maternity leave program has a negative impact on birth outcomes, which is not consistent with the previous studies. Finally, I conduct a placebo test and verify that the violation of the assumption will result in serious differential biases.

1. Introduction

The mechanism of DD is to compare the difference in outcomes after and before the "treatment" for groups that affected by the treatment to the same difference for unaffected groups. The "treatment" is often a social policy, insurance program, or the law, etc. For example,

The outcome Y is modeled by the following equation

$Y_{it} = \beta_0 + \alpha T_{it} + \delta t + \theta_i + \varepsilon_{it}$ (1)

The coefficient of interest is α, and $\alpha = (Y_{s1}-Y_{s0})-(Y_{w1}-Y_{w0})$, supposed that there are s and w two groups.

In order to get a unbiased estimator, we need to assume that the coefficient of t— δ, should be the same between two groups, that is called parallel trend assumption and also the most critical assumption of DD model.

Maternity leave policies are designed to help new working mothers to address the challenges they faced during their pregnancy, therefore improve their and children's health outcome, working environment, job opportunities. It is a very important social welfare policy to promote gender equality, economic growth and lower motherhood penalty. Maternity leave may include paid and unpaid leave.

There are several mechanisms through which maternity leave may affect the birth outcome positively. First, a mother's mental health could play a very important role in the keeping physical well-being during the gestation, and the quality of cares she can provide to a infant after the birth. An appropriate availability of paid time off can significantly help prevent maternal depression and lower stress level. Women are extremely vulnerable to depression or anxiety during pregnancy. Chatterji and Markowitz(2012) used the data from Early Childhood Longitudinal Study to report that the women who has longer maternity leave will suffer from fewer depressive symptoms, and when the leave is paid, get an overall improvement on their mental health. Second, the access to paid maternity leave can affect the birth outcome through receiving more prenatal medical care and vaccinations, extending the breastfeeding rates and duration.

There are serious “access inequalities” for the paid maternity leave as many low-income women lack access to maternity leave. Gornick et al., (2008) shows that even under the current unpaid and compulsory FMLA system, only approximately 60% of United States workers are eligible for benefits. Many low-educated or single mothers are unlikely to financially affordable to take a unpaid maternity leave and continue to work in a unprotected environment that involves long working hours, weight lifting, inevitable noises, etc. A paid maternity leave is very necessary to these low income mothers.

This paper mainly evaluates the efficiency of Difference-in-Differences model in empirical researches by studying the effects of maternity leave on birth outcome (including birthweight, gestation in weeks, the likelihood of low weight birth and premature birth, four substantive indicators and predictors for new infant health conditions) under Temporary Disability Insurance program in five states: California, Hawaii, New Jersey, New York and Rhode Island. This maternity leave came into effects after November 1978, after the enactment of Pregnancy Discrimination Act (PDA). I use the natality data from National Vital Statistics System.

The results show that the access to paid maternity leave reduces the average birthweight of new infant and gestation in weeks, increases the possibilities of the premature and the low birth weight. The results are all statistically significant, however, it is not consistent with previous studies and common senses. For example, Stearns (2015) find that paid leave will reduce the percentage of low birth weight births in state and the effect is strongest for unmarried mothers, who are more likely to be employed and thus benefit from TDI. I conclude that inconsistent outcome is due to no consideration of parallel trend assumption in practice and limitation of DD model. Bertrand et al.,(2004) shows that because of serial correlation, conventional DD standard errors may grossly understate the standard deviation of the estimated treatment effects, leading to serious overestimation of *t*-statistics and significance levels.

This paper is structured as follows: Section 2 discusses the details of TDI and its paid maternity program in USA. Section 3 reviews the related background literature. Section 4 describe the data used to evaluate the policy. Section 5 and presents the summary statistics, and results are mainly revealed in Section 6. Section 7 conduct placebo test and Finally Section 8 provides conclusion.

2. Temporary Disability Insurance (TDI) and maternity leave in USA

Maternity leave mainly refers to a short and temporary time period granted to new mothers to be absent from employment immediately before and after childbirth. It is widely thought that maternity leave is essential to maternal health and child health and development and also very important to economic efficiency and gender equality. The Family Medical Leave Act (FMLA), passed in 1993, is the main legislation that direct the maternity leave in USA. Under FMLA, employers are required to give 12 weeks of unpaid leave for qualifying reasons, including the birth and adoption of a child. Due to the FMLA requirement, as well as the voluntary offer of unpaid maternal leave by private businesses, 60 % workers reported that they could take unpaid leave for the birth of a child in a survey by the United States Social Security Administration.

However, comparing to other developed countries, United States is the only country without a nationwide paid maternity leave among the Organization for Economic Cooperation and Development (OECD) countries. California was the first state in the United States to implement a paid family leave (PFL) program in 2004. Until now, only three states (California, New Jersey and Rhode Island) provide such kind of paid leave for new mothers, although the time period is short (four to six weeks) and coverage is limited. So on the one hand, in general new mothers can access to “paid leave” through using sick leave and vacation, on the other hand, some states have had eligible workers enroll in some state insurance programs, most commonly, disability insurance(DI) program, and pregnancy is defined as a kind of disability. It compensates new mother for economic loss (past and future), reimbursement or payment of medical and life expenses (functioning in this case as a form of health insurance).

In 1978, the U.S. Congress enacted the Pregnancy Discrimination Act (P.L. 95-555) in order to amend the sex discrimination section of the Civil Rights Act of 1964. The act expanded the coverage of the sex discrimination to the pregnancy discrimination—the one "on the basis of pregnancy, childbirth, or related medical conditions." so If an employee is temporarily unable to perform her job due to pregnancy, the employer must treat her the same as any other temporarily disabled employee, then the states with Temporary Disability Insurance are required to provide paid leaves to a pregnant women immediately before and after the birth or adoption.

According to the United States Social Security Administration, Temporary Disability Insurance, also referred to as cash sickness benefits, is defined as a kind of partial compensation to provides to workers for loss of wages and a short period of leave caused by temporary non-occupational disability. Until now, only five States (California, Hawaii, New Jersey, New York and Rhode Island), Puerto Rico, and the railroad industry have temporary disability insurance laws. Before November 1978, the TDI did not generally cover the pregnancy as a kind of disability, but the complications caused by the pregnancy discrimination, then after the enactment of Pregnancy Discrimination Act 1978, the five states are required to extend all maternity benefits to all eligible pregnant women. The passage of the PDA was regard as a major factor in encouraging more new mother participate in the labor force and raise their wages and welfare in that it required employers to provide paid sick leave, health insurance, and TDI benefits long denied them.

There are some important reasons for studying the effects of statewide paid maternity leave. First, paid maternity leave is still inaccessible to most Americans regardless of that young women with young children consist of a significant part of the labor force in the United States today. Pregnancy Discrimination Act prohibits the employers from discrimination women on the basis of pregnancy, however, it does not require the employers to provide a compulsory paid maternity leave. The absence of nationwide paid maternity leave make many female employees, especially low income workers, exposed to unfavorable work environments, which is neither good for their and newborn children's health, or their recovery from the gestation. The study of TDI will reveal how sufficient the introduction of a new paid maternity leave will actually affect child outcome, and provides insights to expand paid leave to other states, even in the country.

3. Literature

Difference-in-Differences model takes into account general changes over time that are common to both treatment and control groups without assuming that we have measured all differences between participants and nonparticipants, so It is popular in empirical economics to estimate the effects of certain policy interventions and policy changes that do not affect everybody at the same time and in the same way. Some limitation are gradually found by researchers, especially for the uncertainty of inference. Bertrand et al., (2004) find that many papers that employ Differences-in-Differences estimation (DD) use many years of data and focus on serially correlated outcomes but ignore that the resulting standard errors are inconsistent. And the serial correlation will lead to serious overestimation of *t*-statistics and significance levels. Abadie et al.,(2010) found another uncertainty is not reflected by the standard errors. They doubt the ability of the control group to reproduce the counterfactual outcome trajectory that the affected units would have experienced in the absence of the intervention or event of interest. All these indicate that a naive Difference-in-Differences approach will lead to serious estimated errors and biases.

Recent studies on some plausible exogenous paid maternity leave policies shows ambiguous effects on children. Liu, and Skans (2010) find that, on average, the reform to extend parental leave benefits from 12 to 15 months for Swedish children born after August 1988 had no significant effect on children's scholastic performance. The result is similar to Danzer and Lavy (2013)'s study on a reform of parental leave in Austria. On the other hand, Chatterji and Markowitz (2012) suggest that longer maternity leave from work, both paid and un-paid would help new mother to improve overall health, such as the reduction in the likelihood of severe depression. Rum (1998) investigates paid parental leave in nine European countries over the 1969 through 1994 period substantially improve children's pediatric health, as measured by birth weights and infant or child mortality.

The birth outcome is affected by paid maternity leave policies via various channels. Maternity leaves increase new mother's free time to spend with children and encourage them to take more beneficial ways to care their children, for instance, breast breeding. Huang and Yang (2015) find an increase of 3–5 %for exclusive breastfeeding and an increase of 10–20 % for breastfeeding at several important markers of early infancy after California first implemented a paid family leave (PFL) program in 2004. And paid maternity can also reduce mother's mental stress level due to financial worries and job security. Aizer et al., (2012) find that in-utero exposure to elevated levels of the stress hormone cortisol will negatively affect offspring cognition, health and educational attainment. And Del Bono et al., (2012) 's study about the antenatal parental behavior on birth outcomes among children in the UK and US, shows that mothers' work interruptions of up to two months before birth have a strongly positive effect on birthweight.

4. Data

This study uses data from the National Center for Health Statistics Vital Statistics natality data from 1972 to 1985 to measure the effects of TDI maternity leave on birth outcome, the whole sample size is 39,606,845, The birth record includes main birth data such as infant's birthweight, gestational age, gender, One-Minute Apgar score etc. This dataset also contains the demographic information about the mother's age (<19 years old, 19-24 years old, 25-34 years old, 35-44 years old, > 45 years old, five levels), race (non-hispanic white, black, hispanic, etc.), education background(high school, some college, college degree or more, etc.), marital status, and state and county of residence which is used to separate treatment and control groups.

In order to make the sample size more manageable, I collapse the raw dataset into birth-year/birth-month/state/mother's education/mother's race/mother's age and mother's marital status cells, and finally I got 4,191,086 cells, with an average of 9.5 birth per cell.

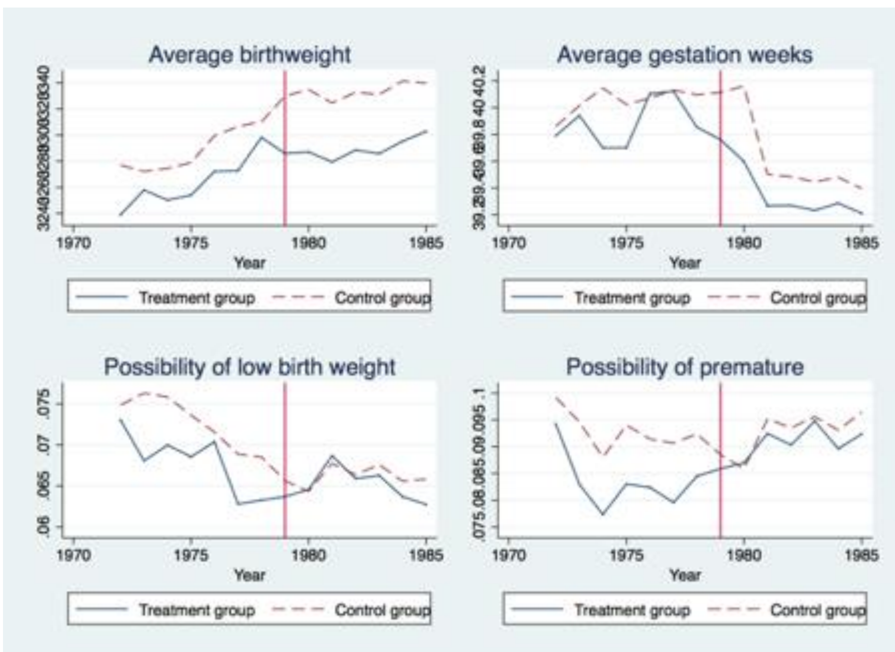

Figure.1. Average change of birth outcomes in two comparison groups

Figure 1 shows the average change in birthweight, low birthweight, gestation and premature births in each TDI state compared to the control groups. Low birthweight is defined as weighting less than 2500 grams and premature is defined as the gestation less than 37 weeks.

5. Estimation model and summary statistics

In order to identify the treatment effect of TDI paid maternity leave on birth outcome, I implement a difference-in-differences framework. I use the fact that only five states with TDIprovided paid maternity leave after November 1978 and thus compare birth outcomes between new infant born in these states and other states before and after that time. The group of these five states are defined as treatment group and other states are called control group.

I estimate following equation:

$$
\begin{aligned}
&Y_{ismy} = \alpha + \beta_1 * POST_{my} + \beta_2 * TREATMENT_s \\
&+\beta_3 * POST_{my} * TREATMENT_s + \gamma' * X_{smy} + \delta_s + \delta_s * timetrend \\
&+\lambda_y + \theta_m + + \eta_{smy} + \varepsilon_{ismy} \qquad (2)
\end{aligned}
$$

Y represents the main birth outcome, such as birthweight, gestation, low birth weight, premature, etc. the POST indicates whether November 1978 or after and the TREATMENT is an indicator for five states with TDI program. The interaction between POST and TREATMENT is the state-month-year indicator and $ß_3$ is the key of coefficient of interest that we want to estimate the effect of TDI on children born in the treatment group. And the vector of X includes state-month-year control variables such as mother's age, race and education etc. and father's age, race and education and gender of child, the population size of the city, etc.

Besides these variables, in order to eliminate the worries that states that are trending up or trending down are more likely to change policy, so I also include group × time dummy variables in the model.

Table 1 presents summary statistics for selected variables in NVSS data for whole country, and by whether are treatment groups, and by whether are before or after TDI.

In the table, we can tell that the whole sample size is 39,606,845, and the size of treatment group (California, Hawaii, New Jersey, New York and Rhode Island) is 7,287,179 and then that of control group is 32,319,666. We also find that the main group of the sample is 19-34 years old, white mothers who hold high school degree, and that is the same both in treatment group and control group.

6. Results on the effects of TDI

Table 2-5 presents the difference in differences estimates of the TDI maternity leave. The first column is the most basic difference-in-differences model regression, which does not include fixed effect control, time trend dummy variable or any control variables. The second column just further includes state fixed effects, month of birth fixed effects and year of birth fixed effects. Then I add the time trend variables which is in the third column. Finally, I include all control

Table 1

Outcomes	Whole Sample			Control state and pre-TDI		Control state and post-TDI		Treatment state and pre-TDI		Treatment state and post-TDI	
	N(whole sample)	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Birth weight in grams	39,606,845	3344.617	668.0117	3329.301	678.2434	3356.992	661.7675	3325.473	688.9177	3346.95	650.0985
5-min Apgar Score	26,520,093	29.7086	37.82905	46.3079	44.26852	24.76981	34.19911	11.88265	15.13191	45.17666	43.9708
Gestation in weeks	39,606,845	44.6299	23.33889	42.43057	31.07363	45.5929	18.32885	49.79219	22.60478	43.36593	15.13299
Low Birth weight(<2500 g)	39,606,845	0.0701125	0.2553365	0.073557	0.2610486	0.0684618	0.2525367	0.071041	0.2568934	0.0667838	0.2496472
Low Apgar Score(<8)	39,606,845	0.0204569	0.141557	0.0042073	0.0042073	0.0338426	0.1808238	0.0048909	0.0697635	0.0186675	0.1353477
Premature(gestation in weeks< 37 week)	39,606,845	0.140167	0.3471603	0.259296	0.4382483	0.0844731	0.2780961	0.0699381	0.2550427	0.0863048	0.2808136
Other Control Variables											
Gender(male =1 , female=0)	39,606,845	0.4873045	0.4998388	0.4871212	0.4998341	0.4875092	0.499844	0.4868735	0.4998278	0.4872145	0.4998366
Mother's education											
Below High school	39,606,845	0.0368637	0.188427	0.0440984	0.2053138	0.0331555	0.1790426	0.0444219	0.2060306	0.0272468	0.1628017
High School	39,606,845	0.5118517	0.4998595	0.5538889	0.4970875	0.5351357	0.498764	0.4143545	0.4926103	0.3500758	0.4769935
Some college	39,606,845	0.1180083	0.322618	0.106384	0.3083284	0.135617	0.3423814	0.0848939	0.2787238	0.0955376	0.293956
college degree or above	39,606,845	0.1340547	0.3407111	0.1141047	0.3179384	0.1543108	0.3612464	0.1032639	0.3043033	0.1225808	0.3279554
Mother's age											
<19 years old	39,606,845	0.1042791	0.3056222	0.1296577	0.3359265	0.0960971	0.2947244	0.0934397	0.2910477	0.0737309	0.2613325
19-24 years old	39,606,845	0.3890628	0.4875376	0.4164356	0.4929676	0.3855444	0.4867237	0.3759692	0.4843722	0.3334709	0.4714532
25-34 years old	39,606,845	0.456218	0.4980795	0.409485	0.4917388	0.4695519	0.4990721	0.4744541	0.4993471	0.5213011	0.4995461
35-44 years old	39,606,845	0.0500504	0.2180491	0.0439357	0.2049521	0.0485087	0.2148386	0.055595	0.229138	0.0710805	0.2569593
>45 years old	39,606,845	0.0003897	0.0197365	0.0004859	0.0220386	0.0002978	0.0172539	0.000542	0.023274	0.0004166	0.0204077

Notes: The units of observation for the summary statistics presented here are state/year/birth month/mother age/mother race/mother education cells. Data is from the universe birth record in the United States for 1972–1985. Post-TDI =Birth occurred in or after November, 1978, Pre-TDI = Birth occurred before November, 1978; Treatment state = Birth occurred in a state that offer TDI program statewide. Control state = Birth occurred in a state without TDI.

variables, for example, the characteristics of mother and father(four dummies for mother's or father's age, three dummies for mother's and father's education levels and three dummies for the race of mother and father, etc.) and the city's characteristics, such as the population size of the city and whether it is in a metropolitan area. Since the likelihood of low birthweight and a premature birth are both between 0 and 1, so I implement the logistic regression instead of OLS that I use on the other two variables. The treatment effect refers to the coefficient of POST*TREATMENT.

From these tables (see Appendix Table 2-5), we can find that the effects of TDI maternity leave on four birth outcome variables —birthweight, the likelihood low birthweight, gestation in weeks and the likelihood of a premature birth are statistically significant. It is worth noting that the coefficients do not change significantly as controls and fixed effects are added into the regression. The results implies that TDI maternity leave have a negative impact on all birth outcomes. For example, the birth weight decrease -3.481 grams and the possibility of low weight birth (<2500 g) increases 3.93% on average after the implement of TDI maternity leave after 1978.

Furthermore, I replace the POST variable with yearly dummies, conduct the same DD estimation on each year, and get the annual DD coefficients. Then I plot these coefficients from DD model, and each point represents the coefficient on the interaction between an indicator for a birth in a treatment group, and an indicator for the year on the x-axis (Figure 2).

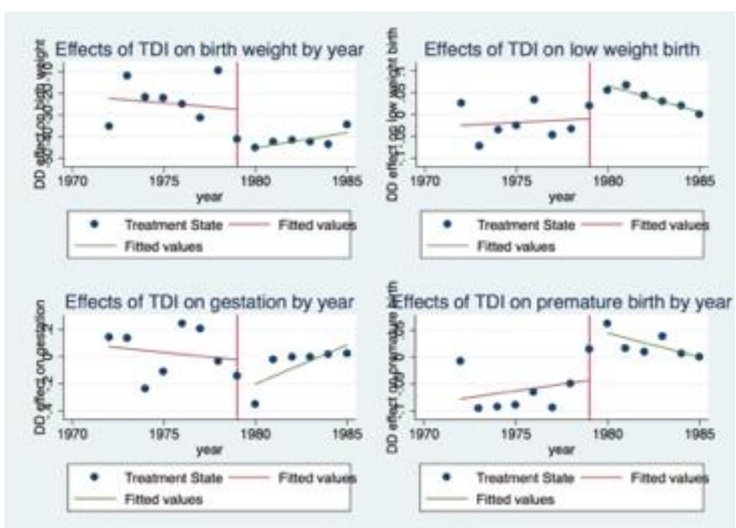

Figure.2. DD Effects by year

The graphs clearly suggest there is a negative effect on birth weight and gestation weeks, and a increase of the likelihood of a low birth weight and premature birth after TDI 1978, which means TDI maternity program do not improve birth outcome in these five states, but deteriorate the health of new infants, even if we have included a time trend variables to get rid of the situation that these states are trending down. It is not consistent with our expectations and the results of previous researches.

One popular explanation for this phenomenon is access inequality: only some parts of whole sample have been affected by the policy. For example, Han et al., (2009) pointed out that the maternity leave policies in U.S only affect mothers possess college and above degrees, married and have stable work. And it is further evidenced by Maya (2011)'s study on the effects of FMLA (the nationwide unpaid maternity leave), in which the effects on births by these women tend to be larger and more statistically significant. Then I divide new mothers into two panels: the young mothers from white and college-and-above-educated-level_ family and other mothers, and conduct the same analysis on them respectively, the result is presented in Table 6. The effects on the birthweight of births by college and above educated mother seems more positive than the births by less educated and mother, however, they are not statistically significant in the sub-group.

Table 6				
Young mothers from white and high-edu families				
	Birthweight	Low Birthweight	Gestation	Premature
POST*TREATMENT	0.628	-0.0201	-0.322***	0.0754
	(0.14)	(-0.46)	(-10.77)	(1.92)
Sample Size(N)	447345			
Other mothers				
POST*TREATMENT	-8.417***	0.0173	-0.0139	0.0475***
	(-5.37)	(1.63)	(-1.26)	(5.03)
Sample Size(N)	3743741			
t statistics in parentheses				
* p<0.05, ** p<0.01, *** p<0.001				

Another possible explanation is that Difference-in-Differences model has not sufficiently identified the treatment effect of the TDI. The assumption of DD is that there are the same time trend between treatment and control groups. However, we can notice from Figure 1 that there is a huge downward trend in the gestation weeks and the increase of birthweight slow down for treatment groups before the enactment of TDI, the negative effects may be driven by this differential trend in two groups. This will be further proved in Section 7 Placebo test.

7. Placebo test the weakness of the model

To test the efficiency of DD framework to identify the treatment effect, I conduct a placebo test which mainly identify the trends in both groups in the absence of TDI program and the results are listed in the Table 7. The mechanism of the placebo test is to include an interactions between the indicator for 1977, 1978 and the indicator for the treatment states into the original DD model with time trend. Since we know that there is no treatment effect prior to the November, 1978. The estimated impact for these years should be not significant. If it is statistically significant, this calls into question, some aspect of the identification strategy and/or estimation procedure. The results are listed in Table 7, and all coefficients are very significant and have large standard errors, so we know there are spurious effects prior to TDI and DD fails to identify the treatment effect.

We can tell a significant downward trend in birth outcomes for the treatment states prior the TDI maternity leave, for example, for birth weight of new babies, the DD effect precipitates from -9.495 in 1978 to -41.39 in 1979 when TD came into effects. After several years, the DD effects on all birth outcomes started to rebound, it indicates that TDI indeed started to have a positive impact. The assumption of DD framework does violate in this situation, and this can also explain why the effects of TDI maternity leave on the birth outcome are negative.

One way to solve this problem is to replace DD model with difference-in-differences-in-differences model (DDD) which does not reply on this assumption that same time trends between treatment and control groups.

$$
\begin{aligned}
Y_{smcy} &= \alpha + \beta_1 * POST_{mcy} + \beta_2 * TREATMENT_s \\
&+ \beta_3 * ELIG_{sy} + \beta_4 * POST_{mcy} * TREATMENT_s + \\
&\beta_5 * POST_{mcy} * ELIG_{sy} + \beta_6 * ELIG_{sy} * TREATMENT_s \\
&+ \beta_7 * POST_{mcy} * ELIG_{sy} ** TREATMENT + \\
&\gamma' * X_{smcy} + \delta_s + \delta_s * timetrend \\
&+ \lambda_y + \theta_{mc} + + \eta_{smcy} + \varepsilon_{smcy} \qquad (3)
\end{aligned}
$$

Table7				
Placebo effects in difference-in-difference model				
Outcome	Birthweight(g)	Low Birthweight	Gestation(weeks)	Premature
Treatment state * 1977(placebo1)	3.359	-0.0628***	0.383***	-0.0553**
	(1.27)	(-3.34)	(20.82)	(-3.27)
Treatment state * 1978(placebo2)	7.146**	-0.0520**	0.0882***	-0.0102
	(2.73)	(-2.79)	(4.84)	(-0.62)
Treatment state * 1979(DD effects)	12.03***	-0.0837***	0.377***	-0.0699***
	(4.82)	(-4.75)	(21.68)	(-4.53)
Sample Size(N)	4191086			
t statistics in parentheses				
* p<0.05, ** p<0.01, *** p<0.001				

In Equation (3), where $ELIG_{sy}$ is an indicator whether the mother's state and year of birth place is into the likely eligible group. Rossin (2010) uses the likelihood of employment in a firm with 50 or more employees above the median to identify eligible and ineligible group.

Besides, the trends of birthweight and gestation are quite different in different parts of the country during this time period, and as we proved above, using the rest states of the country as control group simply and directly will make the assumptions of DD analysis unlikely to hold. It might be not reasonable to think that states with TDI should be compared to all other states in U.S. and we need to select a better control group. Abadie et al.'s (2010) implements a kind of synthetic control method by matching based on a set of characteristics. And Stearns (2015) identifies all possible control groups based on statewide percentage of low birthweight births, percentage of women in each age group, in each race group, and uses a weighted average of these states to create a new synthetic control group and gets positive and significant effects. This would also help to improve the efficiency of DD.

Another shortness of this paper is the insufficiency of controls variables. Because NVSS natality data just includes birth data, the basic demographic information. However, the effect of maternity leave are significantly impacted by the factors beyond family unit. For example, only employed women could be access to the TDI maternity leave, however, we cannot observe whether a woman is employed at the time of birth in this dataset, so we need more labor data such as the employment rate of women and if the woman works for a public or private sector.

8. Conclusion

The lack of nationwide paid maternity leave in U.S. makes many low-income mothers have to make the choice between their self and children's health and job & financial security. The paid maternity policy is very essential to women's career development, physical and mental health, and overall well-being as well as the birth outcome. The study the effects of TDI— a statewide paid maternity leave would help us to see the impacts of the similar paid leave extended to the rest of the U.S. today.

I apply a Difference-in-Differences method and consider various outcomes, and find that TDI paid maternity leave have a significant negative impact on birth outcomes. It looks plausible, but is a biased estimator of the treatment effect of TDI. In order to identify the bias, I try several methods. First, I think paid maternity leave may have increased disparities in early childhood health between children from different social economic background, so I conduct sub-sample analysis on children of college-educated and married mother, and these less-educated and single mother, but do not find a very salient difference. Then the average change of dependent variables between two comparison groups suggests that there might be a neglect of the parallel trend assumption. So I do a placebo test to check the robustness of the original DD model, and finally shows that the DD model fail to estimate the treatment effects. The strong significance level derives from big sample and the overestimation of DD model, the negative effects of TDI mainly are due to a huge downward trend existing in the treatment states.

When implementing DD framework, researchers often select the treatment and the control groups just on the basis of subjective measures of affinity between affected and unaffected units, for example, the states with the policy and the states without it. The comparative case studies typically employ data on a sample of disaggregated units and inferential techniques that measure only uncertainty about the aggregate values of the data in the population.

My study suggests that a naive DD approach would result in estimation errors. The conventional DD model has its limitation: it may overestimate t-statistics and significance levels because of serial correlation, and its standard errors may understate the standard deviation of the estimated treatment effects. I hope that my study can provide some motivation for the empirical researchers who estimate DD models to more carefully examine the assumptions of the model, and more synthetically define the comparison groups, and even examine residuals as well as perform simple tests of serial correlation.

Appendix:

Table2				
Effects of TDI maternity leave on birth weight(grams)				
	DD	DD with fixed effects	DD with time trend	DD with all controls
POST	42.69***	2.752	13.92***	6.882***
	(57.94)	(0.65)	(9.85)	(4.95)
TREATMENT	-21.02***	2.948	-9996.4	2886.8
	(-18.52)	0.00	(-0.20)	(0.06)
POST*TREATMENT	-22.94***	-11.81***	-12.61***	-3.481
	(-16.55)	(-7.95)	(-4.64)	(-1.31)
Sample Size(N)	4191086			
t statistics in parentheses				
* p<0.05, ** p<0.01, *** p<0.001				
Notes: The results in the table 2 is the difference-in-difference effects on birth weight. The units of analysis are state/year/month/mother's education/race/age cells. Controls include: (1)maternal characteristics: four dummies for mother's age, three dummies for mother's education, three dummies for mother's race; same parental characteristics; the gender of child;(2)the population size of the city; the dummy for metropolis.				

Table3				

Effects of TDI maternity leave on low birth weight				
	DD	DD with fixed effects	DD with time trend	DD with all controls
POST	0.0934***	0.0249	-0.0333***	-0.00864
	(-18.86)	(0.86)	(-3.43)	(-0.88)
TREATMENT	-0.0851***	-0.129**	-24.32	24.42
	(-10.55)	(-2.87)	(-0.57)	(-0.57)
POST*TREATMENT	0.0662***	0.0294**	0.0634***	0.0393*
	(6.69)	(-2.85)	(-3.35)	(2.06)
Sample Size(N)	4191086			
t statistics in parentheses				
* p<0.05, ** p<0.01, *** p<0.001				
Notes: The results in the table 2 is the difference-in-difference effects on the likelihood of low weight birth. The units of analysis are state/year/month/mother's education/race/age cells. Controls include:(1)maternal characteristics: four dummies for mother's age, three dummies for mother's education, three dummies for mother's race; same parental characteristics; the gender of child;(2)the population size of the city; the dummy for metropolis.				

Table4				
Effects of TDI maternity leave on gestation weeks				
	DD	DD with fixed effects	DD with time trend	DD with all controls
POST	-0.446***	-0.0482	0.0273**	0.00236
	(-74.72)	(-1.64)	(2.77)	(-0.24)
TREATMENT	-0.164***	0.0737	269.7	281.5
	(-17.38)	0.00	(0.76)	(0.79)
POST*TREATMENT	-0.0838***	-0.0488***	-0.250***	-0.226***

	(-7.99)	(-4.72)	(-13.22)	(-11.97)
Sample Size(N)	4191086			
t statistics in parentheses				
* p<0.05, ** p<0.01, *** p<0.001				
Notes: The results in the table 2 is the difference-in-difference effects on mother's gestation week. The units of analysis are state/year/month/mother's education/race/age cells. Controls include:(1)maternal characteristics: four dummies for mother's age, three dummies for mother's education, three dummies for mother's race; same parental characteristics; the gender of child;(2)the population size of the city; the dummy for metropolis.				

Table5				
Effects of TDI maternity leave on premature				
	DD	DD with fixed effects	DD with time trend	DD with all controls
POST	0.00740	0.0345	-0.0277**	-0.00318
	(1.69)	(1.37)	(-3.26)	(-0.37)
TREATMENT	-0.118***	-0.0855*	-70.78*	-20.41
	(-16.27)	(-2.29)	(-1.99)	(-0.57)
POST*TREATMENT	0.0873***	0.0621***	0.0960***	0.0730***
	(9.94)	(6.77)	(5.76)	(4.35)
Sample Size(N)	4191086			
t statistics in parentheses				
* p<0.05, ** p<0.01, *** p<0.001				

Notes: The results in the table 2 is the difference-in-difference effects on the likelihood of the premature birth. The units of analysis are state/year/month/mother's education/race/age cells. Controls include:(1)maternal characteristics: four dummies for mother's age, three dummies for mother's education, three dummies for mother's race; same parental characteristics; the gender of child;(2)the population size of the city; the dummy for metropolis.

Reference

Abadie, A., Diamond, A.,and Hainmueller,J.(2010), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association*, 105, 493-505

Aizer, A., Stroud, L., and Buka, S. (2012), "Maternal Stress and Child Outcomes: Evidence from Siblings", Working Paper W18422, NBER.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004), "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics,* 119 (1), 249–275.

Chatterji, P., and Markowitz, S.(2008), "Family Leave after Childbirth and the Health of New Mothers," Working Paper W14156, NBER.

Danzer, N., and Lavy, V.(2013), "Parental Leave and Children's Schooling Outcomes: Quasi-experimental Evidence from a Large Parental Leave Reform, " Working Paper W19452 , NBER.

Del Bono, E., Ermisch, J., and Francesconi, M. (2012), "Intrafamily resource allocations: a dynamic structural model of birth weight." *Journal of Labor Economics,* 30 (3), 657–706.

Gault,B., Hartmann, H., Hegewisch,A ., Milli,J.,and Reichlin,L.(2014), "Paid Parental Leave in the United States:What the data tell us about access, usage, and economic and health benefits." Institute for Women's Policy Research (IWPR). ("http://www.iwpr.org/publications/pubs/paid-parental-leave-in-the-united-states-what-the-data-tell-us-about-access-usage-and-economic-and-health-benefits/").

Gornick, J., Ray, R., Schmitt. J(2008), "Parental Leave in 21 Countries: Assessing Generosity and Gender Equality." Center for Economic and Policy Research. ("http://www.lisdatacenter.org/wp-content/uploads/parent-leave-report1.pdf").

Han, W.-J., Ruhm, C., and Waldfogel, J. (2009), "Parental leave polices and parents' employment and leave-taking." *Journal of Policy Analysis and Management 28 (1),* 29–54.

Huang, R., and Yang, M.(2015), "Paid maternity leave and breastfeeding practice before and after California's implementation of the nation's first paid family leave program." *Economics and Human Biology*, 16, 45-59.

NCHS's Vital Statistics Natality Birth Data—1968–2006, 2010. National Bureau of Economic Research. http://www.nber.org/data/births.html (accessed Dec 16, 2015).

Rossin, M.(2011), "The effects of maternity leave on children's birth and infant health outcomes in the United States." *Journal of Health Economics*, 30, 221-239.

Stearns, J.(2015), "The effects of paid maternity leave: Evidence from Temporary Disability Insurance." *Journal of Health Economics*, 43, 85-102.

U.S. Social Security Administration, "Social Security Programs in the United States-Temporary Disability Insurance". https://www.ssa.gov/policy/docs/progdesc/sspus/tempdib.pdf.

Wooldridge, J. M., and Imbens, G. W. (2008), "Recent Developments in the Econometrics of Program Evaluation," Working Paper W14251, NBER.