University of Duisburg-Essen
Faculty of Business Administration and Economics
Chair of Econometrics

Signing Up New Fathers: Do paternity Establishment Initiatives Increase Marriage, Patental Investment, and Child Well-Being?

Causality and Programme Evaluation

Term Paper

Submitted to the Faculty of
Business Administration and Economics
at the
University of Duisburg-Essen

from:

Sunyoung Ji

Reviewer:

Prof. Martin Karlsson, Ph.D, Chair of Health Economics
TA Nikolaos Prodromidis. M.Sc., Chair of Health Economics

Deadline:

01. 10. 2022

Name:	Sunyoung Ji
Matriculation No.:	229979
E-Mail:	sunyoung.ji@gmail.com
Study Path:	M.Sc. Econometrics
Semester:	5th

Contents

List of Figures

List of Tables

List of Abbreviations

1 Introduction

In-hospital Voluntary Paternity Establishment (IHVPE) is implemented for children in unmarried families. The expected main effects of IHVPE was improving child well-being and parental investments by making easy to establish paternity. It does require less costs and less time than the legal paternity establishment that includes complicated process with the court system and DNA testing. According to the author Maya Rossin-Slater (2017), IHVPE increases paternity, but decreases marriages. The effects on father involvements and child well-being is seem to close to zero or even negative value, in other words, this policy hardly achieves its purpose. The author mentions the reason is the influence of IHVPE to parental marriages. Fathers who would get married with child mother in the absence of IHVPE decide to be not married. Thereby, children could not be provided with private health insurance by their biological father's employees, and father would spend less time with their children and provide less support as they live separately from their children. To analyze these treatments effects of IHVPE empirically, the author conduct the quasi-experimental variation in the timing of IHVPE program initiation across stats and years. The analyses of the CPS-CSS data are on the individual level with the statistical model below:

$$Y_{isty} = \beta_0 + \beta_1 IHVPE_{sy} + \gamma' X_{isty} + \phi' C_{st} + \mu_s + \alpha_y + \delta_s y + \epsilon_{isty} \qquad (1)$$

for each mother i, in state s, in survey year t, with a youngest child born in year y. Y_{isty} is an outcome of interest, X_{isty} individual maternal and chlid characteristics, and coefficients $\beta_1$is a measure the effect of the existence of IHVPE in the child's state and year of birth on the Y_{isty}. The details of variables are in Appendix A.1. The identifying assumptions for (1) are uncorrelation between the state-year variation in the timing of IHVPE implementation and other unobserved time-varying determinants of the outcomes of interest, and common trend assumption that the treatment and comparison states would have had similar trends in outcomes of interest in the absence of IHVPE introduction. The result that can be derived from this author's study is very interesting. I wanted to recheck that IHVPE is not successful policy by applying a new method to the existed dataset. To replicate Rossin-Slater's study, the causal forests is desirable, first, it allows to deal with high dimensional data. CPS-CSS dataset contains total 67-dimensional space, and in each replication, the number of variables is from 28 to 29 including covariate, outcomes, cluster and treatment variables. Moreover, the causal forests provide heterogeneous treatment effects: which/How much subgroups vary treatments effects. Lastly, under some assumptions, causal forests offer a pointwise consistent estimate and accurate asymptotic variance and have asymptotically Gaussian and centered sampling distribution which is informative on the statistical inference. The paper proceeds as follows. Chapter 2 introduces CPS-CSS data and the use of causal forests as an empirical methods. Chapter 3 shows steps for building a replication model, main results and evaluation of the model.

2 Data: Current Population Survey Child Support Supplements (CPS-CSS)

Rossin-Slater use the biannual March/April matched CPS-CSS to analyze the effects of IHVPE on parental marriage, birth father's involvements and child well-being. The dataset consists of 48,119 moms that were surveyed both in March annual demographic file and in the monthly April CPS from 1994 to 2008. For the replication, I used the same subsets with original paper. Variable details are given in Appendix 1.

3 Replication with Causal Forests

3.1 Introduction of the Causal Forests

To estimate causal effects of IHVPE, I apply the causal forests which are the machine learning method that captures heterogeneous treatment effects. In the causal forests, each regression tree is created by first drawing a random subsample of the data. The individuals in the subsample are then split into subgroups ("leaves") based on covariate values, where the splits are defined by minimizing some loss function such as the root mean squared error. The process is then repeated and averaged over B trees, resulting in random forest that can be used to predict Y_i for a particular covariate combination $X_i = x$. The estimation with the causal forests proceeds with `grf (generalized random forests)` package in R. `grf` provides non-parametric methods for heterogeneous treatment effects estimation for forest-based statistical estimation and inference (Tibshirani, et al.).

3.2 Identification Assumptions

This replication with causal forests estimation assumes below:

- Conditional unconfoundedness: $(Y_i^1, Y_i^0) \perp D_i \mid X_i$

- Stable Unit Treatment Value Assumption (SUTVA): $Y_i = Y_i^0 + D_i(Y_i^1 - Y_i^0)$

- Overlap assumption: for all $x \in supp(X_i), 0 < P(D_i = 1 \mid X_i = x) < 1$, $P(D_i = 1 \mid X_i = x) = e(x)$

- Exogeneity of covariates

To satisfy conditional unconfoundedness, I create selection index which sorts confounded variables by selecting variables with high importance. More detail is mentioned in Chapter 3.3.2. step4. Overlap assumption is satisfied by positive probability of treatment for each X_i. It can be conformed by the histogram of $e(x) = \mathbb{E}[W_i \mid X = x]$ which is provided in Chapter 3.5.2

3.3 Replication

3.3.1 Variables

The estimation of treatment effects of IHVPE in this paper consists of 4 replications. Each replication estimates treatment effects (W_i) on parental marriages, child private health insurance in all parents groups, child private health insurance in families that do not live with child supports from biological fathers respectively.

Table 1: Treatments Variables in Each Replications

Replications	**Treatment Variables(Y_i)**
Rep 1	mother who is married to child's father
Rep 2	child who has private health insurance coverage
Rep 3	child who has private health insurance coverage
Rep 4	father who covered any gifts, clothes, food, childcare, or medical expenses

End of table.

It is assumed that the moms in the same state could arbitrary correlated within the a state, since the years of IHVPE initiation are vary between different states. Thus, state FIPS is assigned as a cluster in the replication. When it comes to the covariates, it is devided into 4 groups:

Table 2: Treatments Variables in Each Replications

CovariatesX_i	**Variables**
Households controls	mother's age, education, race, total number of children in a household, child's gender, age
States controls	unemployment rate, poverty rate, minimum wages, percentage of population on medicaid
CS controls	universal wage withholding, license revocation for non-payment, joint custody law, log total expenditures on child support enforcement
welfare controls	AFDC waiver, TANF, EITC

End of table.

The treatment variable indicating implication of IHVPE is a dummy variable from 1 to 0.

3.3.2 Replication Steps

- **Step 1: Estimate** $m(x) = \mathbb{E}[Y_i \mid X = x]$ **and** $e(x) = \mathbb{E}[W_i \mid X = x]$
  I train regression forests with outcome variables and the treatment in each replications to

make out-of-bag estimates of $m(x)$ and $e(x)$ to grow the causal forests through

$$\hat{\tau} = \frac{\sum_{i=1}^{N} \alpha_i(x)(Y_i - \hat{e}^{(-i)}(X_i))}{\sum_{i=1}^{N} \alpha_i(x)(W_i - \hat{e}^{(-i)}(X_i))^2}$$

, where $\alpha_i(x)$ is a data-adaptive kernel that measures how often individual training example falls in the same leaf as the the test points x (Athey and Wagger, 2019).

- **Step 2: Grow a causal forest with $m(x)$ and $e(x)$ with tuning parameters**
  I tune all possible parameters by cross-validation with the tuning option `tune_parameters` of `grf`.

- **Step 3: Train the causal forests on all covariates**

- **Step 4: Train the new causal forests with important covariate variables**
  Through the random forests, I check the importance of all variables, then make the selection index including only variables with bigger importance than the mean of overall importance. By training the causal forests with only important variables in low signal situations, I can eliminate features that may be confounded.

- **Step 5: Estimate Average Treatment effects (ATE) of the causal forests**

3.4 Results

Table 3: Results of Replications

.	**Rep 1**	**Rep 2**	**Rep 3**	**Rep 4**
Rossin-Slater	-0.0281 (0.0088)	-0.0263 (0.0101)	.	0.0161 (0.0290)
ATE	-0.0148 (0.0064)	-0.0151 (0.0040)	-0.0281 (0.0143)	0.0389 (0.0196)
RATE	-0.0185 (0.0073)	0.0077 (0.0044)	0.0137 (0.0214)	0.0418 (0.0127)

End of table.

In table 3, Rep1, Rep2, Rep3 and Rep4 are replications to estimate treatment effects of IHVPE on marriage, private health insurance for children in all households, private health insurance for children who are not living with their biological father and child support from birth father to children who are not living with their birth father respectively. The first row shows coefficients of treatment effects of IHVPE on each outcome of interest which estimate by Rossin-Slater. The second and third rows illustrates average treatment effect and rank-weighted average treatment effects that offered by replication. The rank-weighted average treatment effect is a weighted sum of targeting operator characteristics. The Targeting Operating Characteristic (TOC) is a curve comparing the benefit of treating only a certain fraction (q) of units to the overall average treatment effect. A weighted sum of TOC identifies prioritization rules that effectively targets treatment (Tibshirani, et al.). In summary, the results from replications reflects that the treatment effects of IHVPE on marriage and child private health insurance are less significant than the results from Rossin-Slater's paper. On the other hand, Replication 4 shows effect on any child support from birth dad increases after treatment(IHVPE).

- Rank-Weighted Average Treatment Effects (RATE)

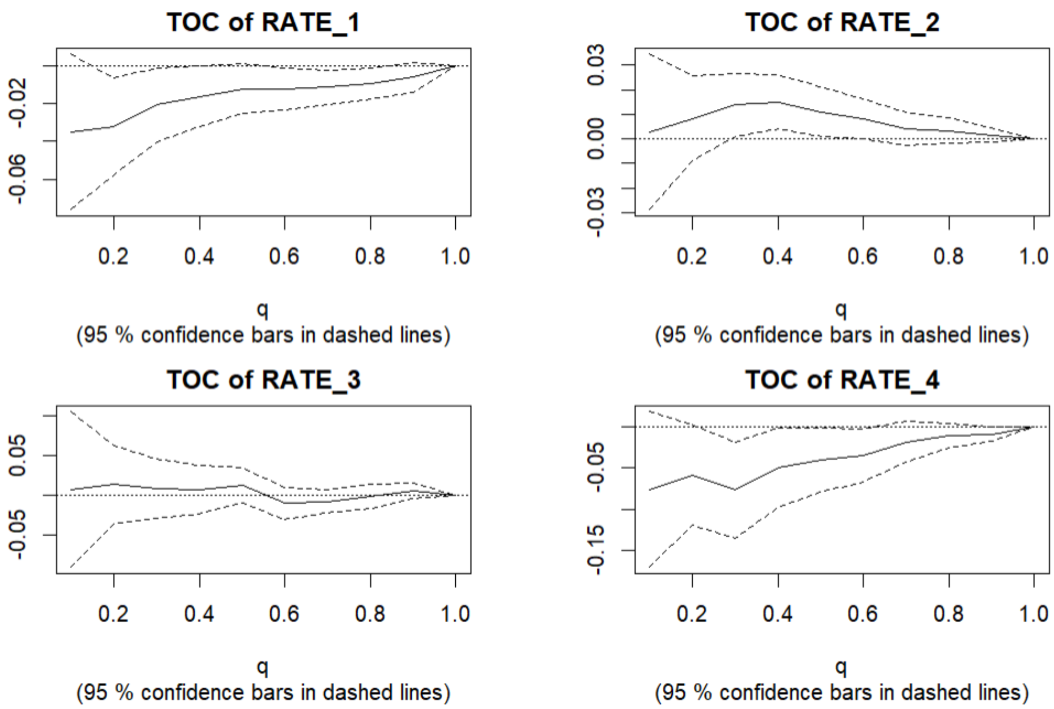

Figure 1: Targeting Operator Characteristics

Table 4 introduces a measure of the quality of the estimates of treatment heterogeneity, $D_i = (\hat{\tau}^{(-i)}(X_i) - \bar{\tau})(W_i - \hat{e}^{(-i)}(X_i))$, where $\bar{\tau}$ is the average of the treatment effect estimates. If estimates equals to 1, the mean of forest prediction is accurate. In Table 5, the estimates are coefficients of the differential forest prediction. It implies that the heterogeneity estimates from the forests are well calibrated, if it is 1 as estimates in Table 4. The p-value of differential forest prediction in Table 5 acts as omnibus test for the existence of heterogeneity. The p-value which is greater than 0 means there is heterogeneity (Tibshirani, et al.).

3.5 Model Evaluation

3.5.1 Evaluation of Treatment Heterogeneity

The best linear predictor method(Chernozhukov, et al., 2018)fits CATE as a linear function of the out-of-bag causal forest estimates(Athey and Wager, 2019). `grf` provies `test_calibration` function that computes the best linear fit of the target estimates using the forest prediction and the mean forest prediction(Tibshirani, et al.).

As Table 4 show, the estimates quite close to 1, especially Rep2 and Rep4. Thus, the causal forest models in the replications predict good the average of the treatment effect estimates. In Table 5, Though estimates are far from 1, except for Rep2, the p-values of Rep2, Rep3 and Rep4 are greater than 0, therefore, I can reject the null hypothesis, that is, there is heterogeneity.

Table 4: Best Linear Predictor Method 1

Replications	estimate	std.err	t.value	p.value
Rep1	0.8255	0.5309	1.5548	0.0600
Rep2	1.0608	0.9583	1.1070	0.1342
Rep3	1.5048	0.4433	3.3944	0.0003
Rep4	1.0505	0.2989	3.5149	0.0002

End of table.

Table 5: Best Linear Predictor Method 2

Replications	estimate	std.err	t.value	p.value
Rep1	1.1624	0.7811	1.4881	0.0684
Rep2	-0.0424	0.4609	-0.0919	0.5366
Rep3	-10.7970	2.2546	-4.7888	1.0000
Rep4	-10.2456	1.9367	-5.2903	1.0000

End of table.

3.5.2 Evaluation of Overlap

The overlap assumption requires a positive probability of treatment for each X_i. None of the estimated propensity scores $e(x)$ should be one or zero. $e(x)$ in each replication has the positive range of values, however, Not a few values are shown to be close to 1. This part will need to be improved for a more robust model in the future.

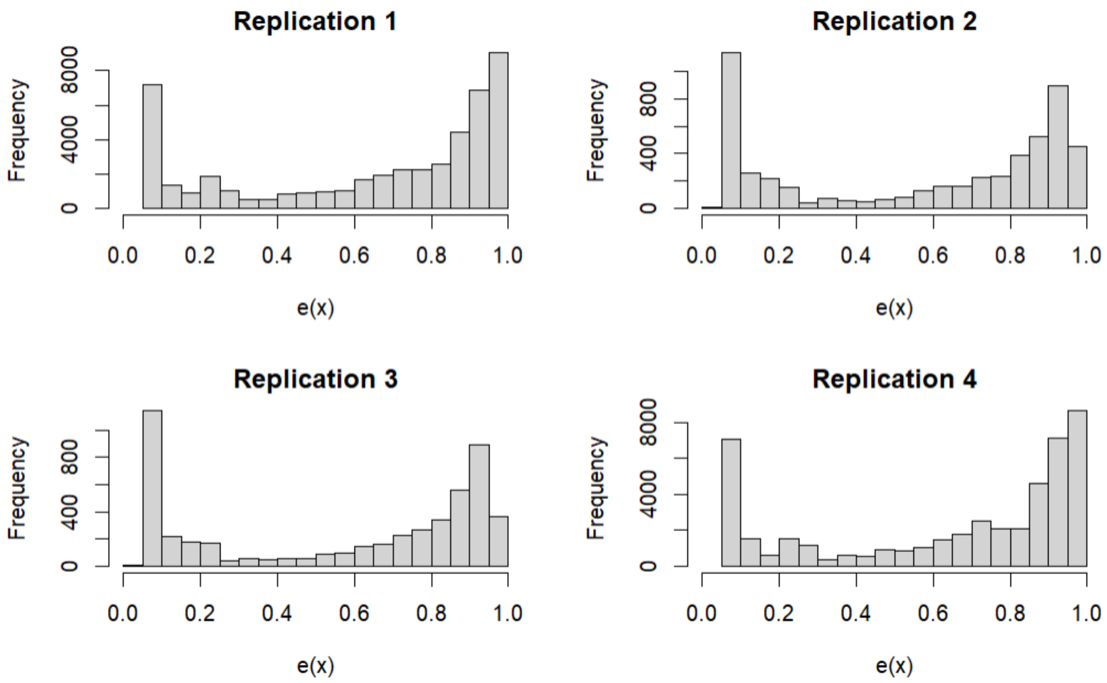

Figure 2: Histogram of e(x)

4 Conclusion

The results of Replications can be summarized as follows. 1) The impact of IHVPE on marriage reduction and child private health insurance is less. 2) The impact of IHVPE on any child support provided by the birth father has increased significantly from 0.0161 in Rossin-Slater's paper to 0.0389. In other words, IHVEPE encourages birth fathers to invest more in their children by improving the relationship between fathers and children. In contrast, there is a limit to IHVPE's influence because most of the private health insurance is provided by employees to their legal families. Despite the imperfections of the current model, IHVPE is not large, but I think it works to some extent.

Refrences

Athey, Susan Wager, Stefan. (2019). Estimating Treatment Effects with Causal Forests: An Application. Observational Studies. 5. 37-51. 10.1353/obs.2019.0001.

Rossin-Slater, Maya. doi: 2017. "Signing Up New Fathers: Do Paternity Establishment Initiatives Increase Marriage, Maya Rossin-Slater, Parental Investment, and Child Well-Being?" American Economic Journal: Applied Economics, 9 (2): 93-130. doi:10.1257/app.20150314

Victor Chernozhukov Vira Semenova, 2018. "Simultaneous inference for Best Linear Predictor of the Conditional Average Treatment Effect and other structural functions," CeMMAP working papers CWP40/18, Centre for Microdata Methods and Practice, Institute for Fiscal Studies. <https://ideas.repec.org/p/ifs/cemmap/40-18.html>

A Appendix

A.1 Description of Relevant Variables

Table A1: Description of Relevant Variables

Variable	Controls	Description
momage1	household controls	mother aged 20 at time of childbirth
momage2	household controls	mother aged 20-24 at time of childbirth
momage3	household controls	mother aged mother aged 25-34 at time of childbirth
momage4	household controls	mother aged 35-44 at time of childbirth
momeduc1	household controls	mother with high school education
momeduc2	household controls	mother with high school degree
momeduc3	household controls	mother with some college
momwhite	household controls	mother is non-Hispanic white
momblack	household controls	mother is non-Hispanic black
momhispanic	household controls	mother is Hispanic
ch_male	household controls	child is male
child_age	household controls	child's age in years
father_outsidehh	household controls	child's age in years
totkids	household controls	total number of kids in the household
unemp_rate_lag1	state controls	last year's unemployment rate
pov_rate_lag1	state controls	last year's poverty rate
state_minwage_lag1	state controls	last year's state-level minimum wage
pct_welf_recp_lag1	state controls	last year's percentage of population receiving welfare
wage_withold	cs controls	universal wage withholding implemented in state/year
new_hires	cs controls	New Hires Directory implemented in state/year
license_revoke	cs controls	License revocation for non-payment implemented in state/year
joint_custody	cs controls	joint custody law implemented in state/year
log_CS_exp_lag	cs controls	log total expenditures on child support enforcement last year
waiver	welfare controls	AFDC waiver implemented in state/year
tanf	welfare controls	TANF implemented in state/year

Continued on the next page.

Table A1: Description of Relevant Variables (continued)

Variable	**Controls**	**Description**
eitc	welfare controls	state EITC program implemented in state/year
state	cluster	state EITC
child_hcovpriv	potential outcomes	child has private health insurance coverage
married_tobiodad	potential outcomes	mother is married to child's father
any_hosp_pat	treatment	IHVPE exists in state and year

End of table.