How does Single Parent Family Affect Children's Human Capital Investment

By Yi Jie Wu, Xiang Jyun Jhang[*]

We estimate the impact of living in the single parent family on the children's future educational attainment by using a comprehensive panel survey data in Taiwan (TEPS). By conducting PDS method, it's found that the experience of living under single parent family will decrease the probability of attaining university degree by percent, and master degree by percents also. We also finds that there is a gap between the probability for two different groups of sample, which we suggest is resulted from the Taiwan Educational Reform in 2000s. Keywords: Single Parent, Labor, Education

American Economic Review Pointers:

- Do not use an "Introduction" heading. Begin your introductory material before the first section heading.

- Avoid style markup (except sparingly for emphasis).

- Avoid using explicit vertical or horizontal space.

- Captions are short and go below figures but above tables.

- The tablenotes or figurenotes environments may be used below tables or figures, respectively, as demonstrated below.

- If you have difficulties with the mathtime package, adjust the package options appropriately for your platform. If you can't get it to work, just remove the package or see our technical support document online (please refer to the author instructions).

- If you are using an appendix, it goes last, after the bibliography. Use regular section headings to make the appendix headings.

- If you are not using an appendix, you may delete the appendix command and sample appendix section heading.

- Either the natbib package or the harvard package may be used with bibtex. To include one of these packages, uncomment the appropriate usepackage command above. Note: you can't use both packages at once or compile-time errors will result.

* Wu: Department of Economics, National Taiwan University, b08302129@ntu.edu.tw; Jhang: Department of Economics, National Taiwan University, b08303024@ntu.edu.tw. We are sincerely grateful for comments and advices from prof. Tzu-Ting Yang

I. Data and Sample

The dataset we use are Taiwan Education Panel Survey (TEPS) and Taiwan Education Panel Survey and Beyond (TEPS-B), in which there are two group of children being surveyed consistently from their schoold age (16-18) to their working age (25-30) and each contains roughly 20,000 samples. The former dataset contains the in-school performance and characteristics of each children by surveying on their parents, teacher of each subject and children themselves. The latter dataset contains the characteristics and working status after the children entered into the labor market, and tracking down the sample for several years. In the whole dataset, there are two main groups of sample, *senior high (SH)* group and *core population (CP)* group. The only difference between these two groups is the born year: the born year for SH group is 1984-1985, but for CP group is 1988-1989. But this unneglected difference leads us to divide them into two groups and perform the same analysis, and it shows us some unexpected result which should be revealed later.

Since our goal is to estimate the causal effect of living in single parent family on children's future educational attainment, we first start on clarifying the condition of **living in single parent family**. As we restricted the period we concerned to high-school age, which is usually the rapidly changing period in life, we define our treatment variable as below:

$$SP_i = \begin{cases} 1, & \text{if child } i \text{ once under SP family in senior high} \\ 0, & \text{o.w} \end{cases}$$

On the other hand, for identifying the children's future educational attainment, we choose three different kinds of outcome variable, which are **university degree, master degree and public university**. The first and second variable are straight-forward, that is whether they obtain the degree when data recoreded. The last one is more about the academic performance measured at the end of their high school age, since the entrance of public university is more competitive than private ones. All these covariates are defined by using the data from TEPS-B, which provides the educational degree and whether the children enter job market after they're above 23-24 years old, so we can identify their final educational attainment.

Finally, the comprehensive dataset also includes the survey data on their parent and teachers, in which the evaluation and observation on the children in their school age are detaily recorded. We consider these dataset as the cure to omitted variable bias, so we look at each teacher's data thoroughly and pick every potential confounding variable that either relates to treatment or outcome variable. Besides, it's worth mentioning that since we want to reduce biasedness, we pay a lot of attention to not include a covariate regarded as **bad control** by us. As the following setup is done, we can move forward to the PDS analysis.

II. Empirical Method

The Post-Double Selection (PDS) method combining causal inference and machine learning is a very powerful and convenient method when we face this kind of huge and comprehensive dataset. Usually, it's the **curse of dimensionality** problem we will encounter when we include more covariates and try to reduce the underlying omitted variable bias; however, the PDS method tackle with this problem with introducing Lasso (Least Absolute Shrinkage and Selection Operator) and regression with multiple stages.

There are two assumptions supporting PDS method: conditional independence assumption and approximate sparsity, and three regression stages in PDS. The first and second stage use Lasso to conduct regression on different dependent variable, that is, treatment variable and outcome variable, which can be formulated as below

$$(1) \qquad \mathrm{SP}_i = \theta_0 + \sum_{i=1}^{n} \theta_j X_i^j + \lambda_\theta \sum_{i=1}^{n} |\theta_j| + \epsilon_i$$

$$(2) \qquad Y_i = \eta_0 + \sum_{i=1}^{n} \eta_j X_i^j + \lambda_\eta \sum_{i=1}^{n} |\eta_j| + \delta_i$$

where the X^j is the covariate we choose from the dataset, and $\{\lambda_\theta, \lambda_\eta\}$ are tuning parameter in the Lasso regression. After the stages, we will have two different sets of selected covariates, denoted as Λ_θ, Λ_η, and their union represents all the potential confounders. So we can conduct the third stage OLS regression to identify the causal effect in the presence of CIA:

$$(3) \qquad Y_i = \beta_0 + \beta_{\mathrm{PDS}} \mathrm{SP}_i + \sum_{j \in \Lambda_\theta \cup \Lambda_\eta} \beta_j X_i^j + \epsilon_i$$

After we follow the same analysis process for each group, SH and CP, we obtain the estimated effect and discover the unexpected result which we will show in next part.

III. Result

Accordingly, we perform the three different kinds of regression model, so as to gradually eliminate the omitted variable bias. The first model is the single variable OLS, which gives us some interpretation of the correlation, the second model is the extension, in which we add some controlled variable related to the student's background; and the last model is the PDS regression, where we put in all the possible confounders and awaits for machine's selection. We should first show our result on the first outcome variable: university degree

(l)2-4(l)5-7	SH (1)	(2)	(3)	(4)	CP/NP (5)	(6)
	University	University	University	University	University	University
SP	-0.0799***	-0.0318***	-0.0136	-0.0989***	-0.0230**	-0.0264**
	(0.000)	(0.008)	(0.402)	(0.000)	(0.022)	(0.041)
female		0.00945			0.000753	
		(0.180)			(0.902)	
hs_private		-0.0762***	-0.0386***		-0.0896***	
		(0.000)	(0.000)		(0.000)	
hs_urban		0.0711***			0.0294***	0.0124
		(0.000)			(0.000)	(0.159)
general_high		0.556***	0.514***		0.635***	
		(0.000)	(0.000)		(0.000)	
paedu		0.101***	0.0739***		0.0829***	0.0330***
		(0.000)	(0.000)		(0.000)	(0.000)
PDS_control			Yes			Yes
N	11132	11050	5230	12576	12386	5766

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

IV. Discussion and Conclusion

References here (manual or bibTeX). If you are using bibTeX, add your bib file name in place of BibFile in the bibliography command.