

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

Yi Jie Wu¹ Xiang Jyun Jhang²

¹Department of Economics
National Taiwan University

²Department of Economics
National Taiwan University

2023/06/12

Quick Recap

What's out research focused on?

- Motivation
 - ▶ **Single parent (SP)** family might fail to provide children with stable environment for learning
 - ▶ Besides, single parent might not give enough mental support to children
- Therefore, we aim to estimate the negative impact on **children's education attainment**

Dataset

Taiwan Education Panel Survey and Beyond, SRDA

- A panel data tracking down two different groups of children across almost 20 years
 - ▶ Senior High group (**SH**): born in 1984-1985
 - ▶ Core/New Population group (**CP/NP**): born in 1988-1989
- A comprehensive dataset **surveying on the children, their parent and teachers**, even after they enter labor market
- Each group contains 20,000 samples

Concerned Variable

What's the outline of our analysis?

- Treatment variable:

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

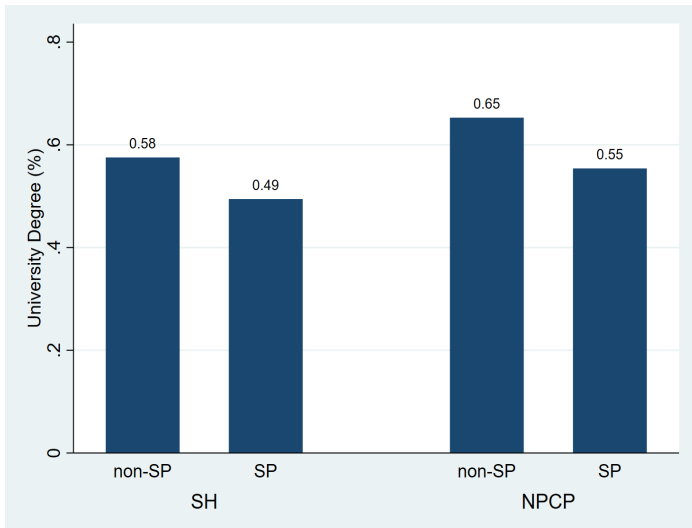
- Outcome variable:

$$Y_i = \{\text{University Degree, Master Degree, Public University}\}$$

- To begin an analysis, it's always good by looking at the graph

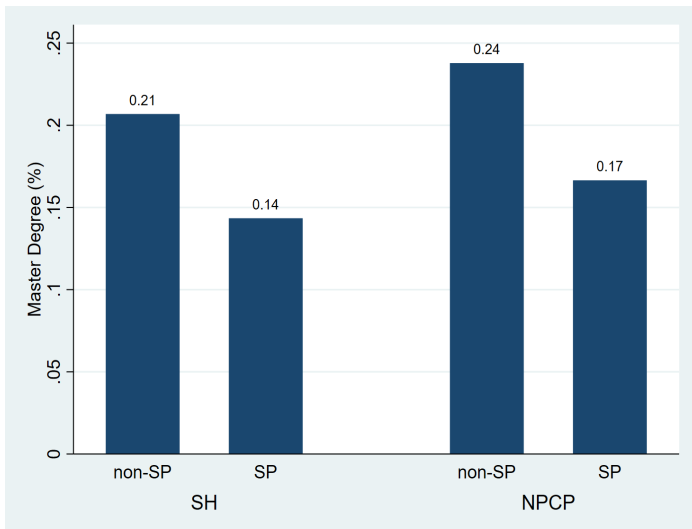
Concerned Variable

Outcome Variable (1): University Degree



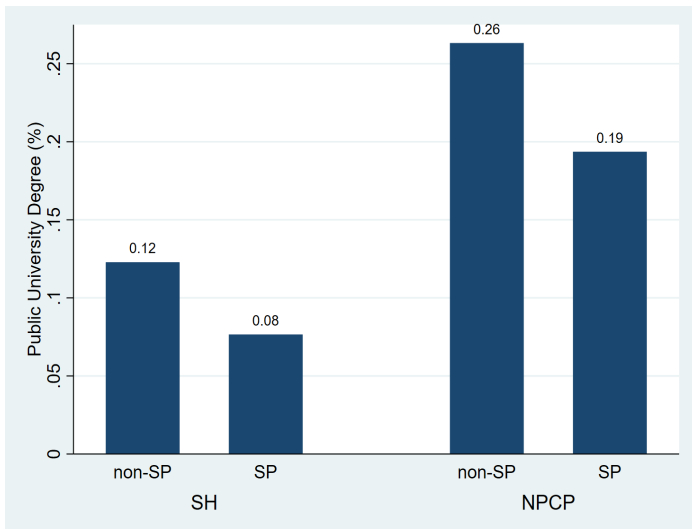
Concerned Variable

Outcome Variable (2): Master Degree



Concerned Variable

Outcome Variable (3): Public University



Concerned Variable

How to pick up potential confounders?

- To perform PDS analysis, We pick an abundance of covariates, including
 - ▶ student's background information
 - ★ gender, living area, private/public school, general/vocational school
 - ▶ parent's education
 - ▶ each teacher's evaluation
 - ▶ etc.

Empirical Specification

Post-Double Selection

- It's important to not include **bad control** in the model, hence we use very few covariates in parent's dataset
- With PDS, we can detect the confounders in roughly **80 selected covariates** and identify the causal relationship

$$Y_i = \beta_0 + \beta_1 SP_i + \bigcup_{j \in A \cup B} \pi_j W_i^j + \epsilon_i$$

where A, B are the Lasso-selected covariates at step 1 & 2 in PDS.

Preliminary Result

Outcome Variable (1): University Degree

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

p-values in parentheses

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

Preliminary Result

Outcome Variable (2): Master Degree

	SH			CP/NP		
	(1)	(2)	(3)	(4)	(5)	(6)
	University	University	University	University	University	University
sp	-0.0753*** (0.000)	-0.0481*** (0.000)	-0.0405** (0.024)	-0.0714*** (0.000)	-0.0404*** (0.000)	-0.0465** (0.015)
female		-0.0981*** (0.000)	-0.0597*** (0.000)		-0.101*** (0.000)	-0.0558*** (0.000)
hs_private		-0.0763*** (0.000)			-0.0555*** (0.000)	
hs_urban		0.0536*** (0.000)			0.0497*** (0.000)	0.0570*** (0.000)
general_high		0.204*** (0.000)	0.156*** (0.000)		0.167*** (0.000)	
paedu		0.156*** (0.000)	0.129*** (0.000)		0.105*** (0.000)	0.0813*** (0.000)
PDS_control			Yes			Yes
N	10438	10362	4848	12173	11986	5600

p-values in parentheses

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

Preliminary Result

Outcome Variable (3): Public University

	SH			CP/NP		
	(1) University	(2) University	(3) University	(4) University	(5) University	(6) University
sp	-0.0331*** (0.006)	-0.00861 (0.450)	0.01000 (0.553)	-0.0696*** (0.000)	-0.0374*** (0.001)	-0.0443** (0.020)
female		-0.00220 (0.757)			-0.0278*** (0.000)	
hs_private		-0.103*** (0.000)	-0.0675*** (0.000)		-0.120*** (0.000)	-0.0860*** (0.000)
hs_urban		0.0751*** (0.000)			0.0841*** (0.000)	0.0968*** (0.000)
general_high		0.229*** (0.000)	0.184*** (0.000)		0.275*** (0.000)	
paedu		0.118*** (0.000)	0.0978*** (0.000)		0.108*** (0.000)	0.0641*** (0.000)
PDS_control			Yes			Yes
N	10577	10503	5027	11518	11353	5576

p-values in parentheses

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

Concluding and Future Prospect

What do PDS tells us?

- There're many unexpected variables PDS selected, like
 - ▶ *"How many years had the teacher taught in school?"*
 - ▶ *"What's the score/performance for this class?"*
 - ▶ *"Does the teacher hold extra class in the weekend?"*
 - ▶ *"How many student's parents the teacher had met with in this semester?"*
- PDS surely help us reduce the OVB in a convenient way
 - ▶ however, missing data problem comes with dimension, leading to significantly decreasing number of samples

Concluding and Future Prospect

What and How can we dive deeper?

- Focus on the difference between two groups (SH vs. CPNP)
 - ▶ high education reform in 2000s
- Explain the negative effect (Mediation Analysis)
 - ▶ accompanying time
 - ▶ economic condition