Do You Remember? Proxy Bias and Recall Bias in Social Mobility Studies: Evidence from TEPS and TEPS-B

Yung-Yu Tsai

Department of Public Administration
National Chengchi University

2018.10.4

Overview

Introduction - Background

Proxy Bias & Recall Bias in Social Mobility Studies

- Social mobility studies focus on the correlation between social class **original** and **destination**.
- Panel surveys are always costly $\rightarrow$ rely on proxy, recall data
- What were your parents' occupations when you were at age 15?

Bias

- Systematical, unrandom
- Would threaten internal validity

$\neq$

Error

- Random
- Would not cause serious problem

- Respondents intentionally or unintentionally revises memories
- Ignorance or amnesia related to certain demographic variables

Introduction - Research Question

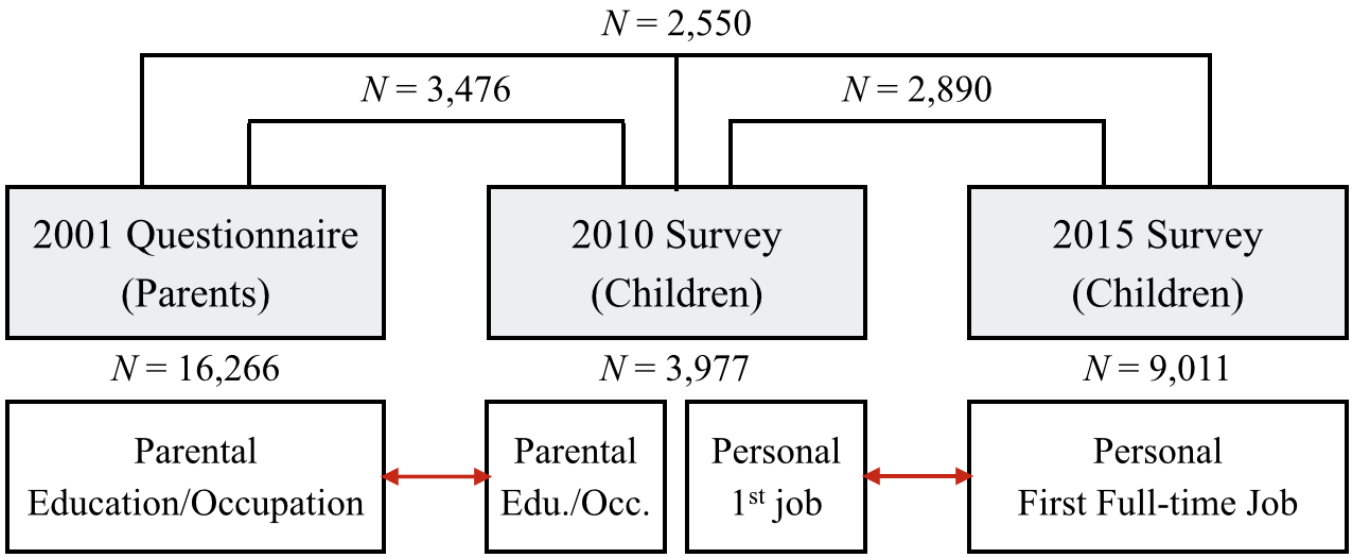
1. Is there any bias in:
   - Children's proxy, recall responses of their **parents' educations**?
   - Children's proxy, recall responses of their **parents' occupations**?
   - Personal recall responses of their **first jobs**?
2. What's the patterns and extents of these biases?
3. How would the above biases affect statistical inference? Is there any way to avoid or reduce these problems?

Research method

Data Taiwan Education Panel Survey (**TEPS**) and its beyond (**TEPS-B**)

Subjects Students who attended grade 11 (senior-high school) in 2001 (1984/1985 cohort)

Method Merge 2001 Questionnaire(parents), 2010 survey(children) & 2015 survey(children), and examine the **consistency** & **correlation** between two different respondents or/and two survey years.



Results - Parental Education

Compare children proxy with parent-report data about parental education

- The ratios of consistency are all above 80%
- Spearman's rho are all significant and greater than 0.85 (strong positive correlation)
- A small gap between self-report data and spouse proxy data

	Father		Mother		Total
	self-report	spouse proxy	self-report	spouse proxy	
Consistency	83.76%	83.83%	86.69%	84.86%	84.84%
Inconsistency	16.24%	16.17%	13.31%	15.14%	15.16%
Spearman's rho	.8791***	.8886***	.8938***	.8598***	-
N=	1,459	1,911	1,893	1,453	6,716

*$p<.05$ **$p<.01$ **$p<.001$

Results - Parental Education

- Most serious bias: parents belong to the group junior colleges
- Children tend to underestimate their parental education
  - **Underestimation**: 8.29%; **Overestimation**: 6.86%
  - Overstate the intergenerational mobility of education or understating parental education's effect on status attainment

Parental answers		Children's answers					
		1	2	3	4	5	Total
1	Junior high or below	94.4%	4.8%	0.1%	0.7%	0.0%	100%
2	Senior high school	11.8%	81.4%	5.8%	0.8%	0.2%	100%
3	Junior colleges	2.7%	15.7%	70.0%	9.8%	1.8%	100%
4	Academic Bachelor	0.9%	2.2%	9.4%	80.2%	7.2%	100%
5	Graduate school	1.8%	4.5%	2.7%	10.8%	80.2%	100%

Concept and measurement of occupation

- Use both International Standard Classification of Occupations (**ISCO**) and The **EGP** Class Scheme to measure "occupation"

ISCO		EGP	
code	category name	code	category name
1000	Managers	I+II	Service class
2000	Professionals		
3000	Technicians and associate professionals	III	Routine non-manual workers
4000	Clerical support workers		
5000	Service and sales workers		
X	Category into other groups	IV	Petty bourgeoisie
6000	Skilled agricultural, forestry and fishery workers	IVc	Farmers
7000	Craft and related trades workers	V+VI	Skilled workers
8000	Plant and machine operators, and assemblers		
9000	Elementary occupations	VIIa	Non-skilled workers
		VIIb	Agricultural labourers

Results - Parental Occupation (measuring in ISCO)

- The ratios of consistency are around $\frac{1}{3}$
- half of the samples exist a serious inconsistency (diff > 1)
- Spearman's rho are all significant but below 0.5 (moderate positive correlation)
- Spouse proxy data are also different from self-report one

	Father		Mother	
	self-report	spouse proxy	self-report	spouse proxy
diff=0 (consistent)	28.15%	26.87%	36.74%	33.48%
diff=1 (inconsistent)	22.66%	21.02%	19.81%	18.42%
diff>1 (inconsistent)	49.19%	52.12%	43.45%	48.09%
Spearman's rho	.4640***	.4747***	.4297***	.3366***
N=	1,112	1,418	1,459	1,102

*$p<.05$ **$p<.01$ **$p<.001$

Results - Parental Occupation (measuring in EGP scheme)

- The ratios of consistency are around 40%
- $\frac{1}{3}$ of the samples exist a serious inconsistency (diff > 1)
- Spearman's rho are all significant but below 0.5 (moderate positive correlation)
- Spouse proxy data are also different from self-report one

	Father		Mother	
	self-report	spouse proxy	self-report	spouse proxy
diff=0 (consistent)	50.90%	48.38%	39.03%	37.39%
diff=1 (inconsistent)	23.14%	23.47%	27.66%	25.64%
diff>1 (inconsistent)	25.96%	28.14%	33.31%	36.98%
Spearman's rho	.4920***	.4685***	.4137***	.3280***
N=	1,275	1,606	1,609	1,217

*$p<.05$ **$p<.01$ **$p<.001$

Results - Parental Occupation (ISCO *versus* EGP)

- Inconsistency is higher when measuring in ISCO than EGP
- Children tend to **underestimate** their parents' occupation
  $\rightarrow$ Overstate the rate of upward mobility
- More common for children to underestimate their mothers' occupation than fathers

	ISCO			EGP Scheme		
	Father	Mother	Total	Father	Mother	Totla
Underestimate	41.26%	48.38%	44.84%	28.81%	46.14%	37.39%
Consistency	27.43%	35.34%	31.41%	49.50%	38.32%	43.96%
Overestimate	31.30%	16.28%	23.74%	21.69%	15.53%	18.64%

Results - Personal first job

2010 Survey

Please recall **all** of the jobs you have engaged in from the time you begin your first job (including part-time and full-time jobs).

2015 Survey

What's your **first full-time job**?
(not mention without asking: the definition of first full-time job is average work hours per week greater than 30 hours)

- **Variable definition**: Capture the first job that its average work hours per week is above 30 hours to represent the first full-time job of a certain subject in 2010 survey
- $\frac{1}{3}$ of samples recognize the same beginning year in two surveys
- Average recall period is 2.5 years in 2010 and 7.5 years in 2015

Results - Personal first job

- The ratios of consistency are 61% (for ISCO) and 80% (for EGP)
- Spearman's rho are above 0.6 (moderate to high correlation)
- The biases are slighter than in parental occupation case

Consistency	ISCO		EGP	
	All samples	Samples with same year	All samples	Samples with same year
diff=0	41.50%	60.70%	66.63%	79.68%
diff=1	23.62%	25.27%	15.05%	12.57%
diff>1	34.89%	14.04%	18.32%	7.75%
Spearman's rho	.3704***	.6515***	.3755***	.6305***
N=	2,511	748	2,511	748

*$p<.05$ **$p<.01$ **$p<.001$

Results - Personal first job

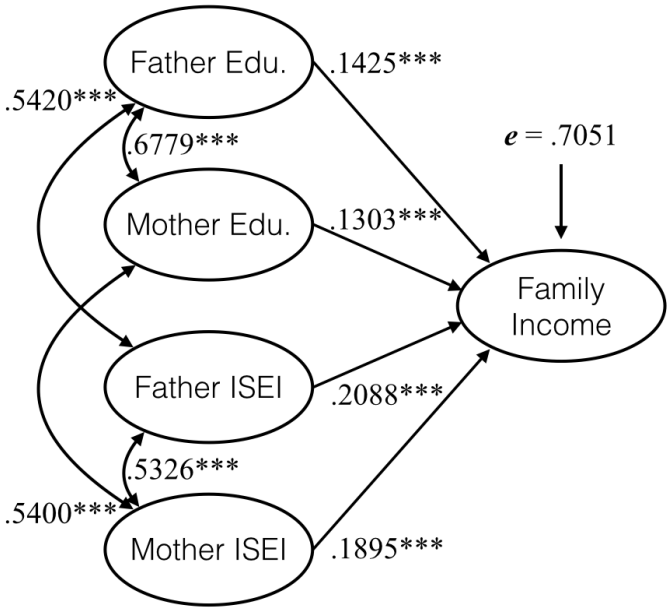
- Compare 2015 answers (recall longer) to 2010 answers (recall shorter)
- The percentage of overestimation is higher than the percentage of underestimation

	ISCO	EGP
Underestimation	14.57%	8.29%
Consistency	60.70%	79.68%
Overestimation	24.73%	12.03%
N=	748	748

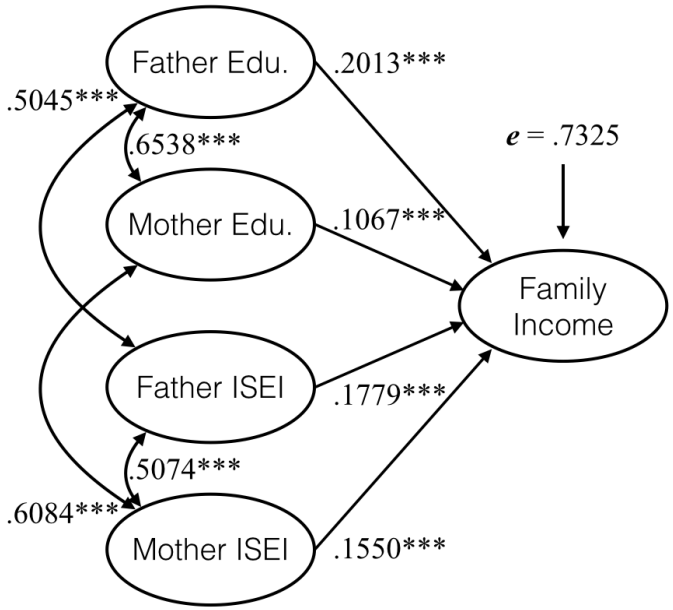
Result - Evaluation of Validity

- Evaluate validity of parent-report data and children proxy data
  - Level of education should be related to occupation status
  - Education and occupation of a couple should be related to each other
  - Education and occupation status should predict family income

Parent-report data



Children proxy data



Conclusion

Parental Education

- Only a **negligible inconsistency** between self-report data and children proxy data
- It's acceptable to use proxy data when measuring simple variables, but researchers should double-check their design of chosen items

Parental Occupation

- A **huge gap** between self-report data and children proxy data
- Children tend to **systematically underestimate** their parental occupation
- The extent and pattern of bias vary from different measurements of occupation and different subjects
- Parent-report data have **better construction validity** than children proxy
- It's better for researchers to get self-report, non-recall data; otherwise they should use **simple classification** to measure variables

Conclusion

Personal first job

- Respondents will **identify different jobs** according to different designs of questionnaires
- Though panel surveys can avoid proxy and recall data, the design of the questionnaire in the past can not always fit the current need
- A small recall bias exist but **reluctantly acceptable**
- Respondents tend to **minimize the gap** between their first job and current class position
- Be careful in questionnaire designing, variables using and result infering