Ability bias in the returns to schooling:

How large it is and why it matters

Petr Čala

Institute of Economic Studies
Charles University, Prague

Introduction

Motivation

- Does ability bias affect the estimation of returns to education?
- Two extensive meta-analyses on the topic (1754 and 293 observations)

Findings

- Average effect of returns to education of around 7%
- Drops by around one percentage point after correcting for publication bias
- Ability matters, and controlling for it in a regression decreases the expected returns to education
- The returns drop even further for twin studies with identical inherent ability (4% to 6%)

What Is Ability Bias

Mincer Equation (Mincer, 1974)

$$\text{Wage} \sim \text{Schooling} + \text{Experience} + \text{Experience}^2$$

- Returns to education: The increase in earnings due to an additional year of schooling
- Ability bias: Distorted estimation of returns to education due to omission of ability (Blackburn & Neumark, 1993)
- Ability correlates with both education and earnings
- Sorting bias: Correlation between ability and education
- How to separate the effect of education from the effect of ability?

Ways To Deal With Ability Bias

- Inclusion of Ability Measures
  - Use cognitive test scores as control variables
  - Separates effect of education from ability
- Instrumental Variables (IV)
  - Find variable correlated with education, not with error term
  - Isolates exogenous variation in education
- Sibling and Twin Studies
  - Compare siblings/twins with different education levels
  - Controls for family and genetic factors
- Other Methods
  - Fixed Effects Models
  - Nonparametric methods

What do we already know?

Study name	AB	AB*	PB	PB*	Method
Psacharopoulos (1994)	.	.	.	.	.
Fleisher et al. (2005)	.	.	.	.	✓
Churchill & Mishra (2018)	.	.	✓	✓	✓
Psacharopoulos & Patrinos (2018)	.	.	.	.	.
Patrinos & Psacharopoulos (2020)	.	.	.	.	.
Cui & Martins (2021)	.	.	✓	✓	✓
Iwasaki & Ma (2021)	.	.	✓	.	✓
Ma & Iwasaki (2021)	.	.	✓	✓	✓
Wincenciak et al. (2022)	✓	✓	.	.	✓
Horie & Iwasaki (2023)	.	.	✓	.	.
Number of studies:	1	1	5	3	6
Percentage of studies:	10%	10%	50%	30%	60%

My contribution

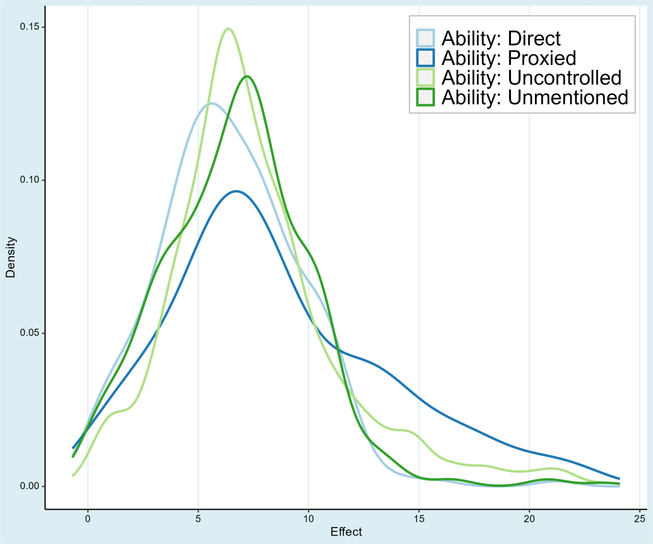
- A large meta-analysis of 1754 estimates of returns to education over 115 studies
- Correct for publication bias, observe heterogeneity
- Observe the isolated effect of ability
- Conduct a whole another meta-analysis comprised of twin studies (293 observations)
- Fully automate the whole analysis process

Different Approach to Ability

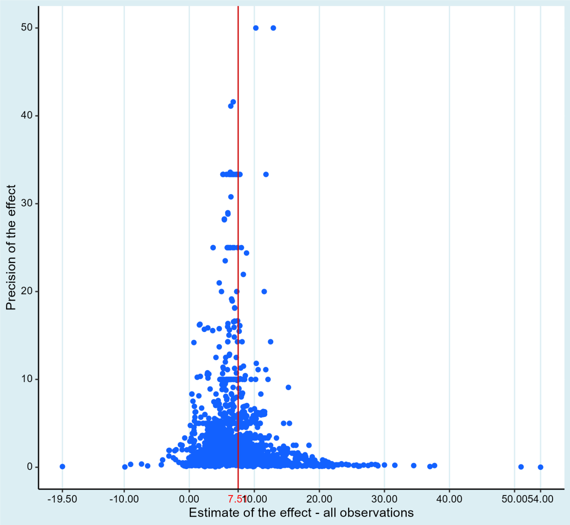
Four ways to address ability:

- Directly - using cognitive test scores or proxies thereof
- Indirectly - using instrumental variables or other methods
- Verbally - acknowledging the issue
- Not at all - ignoring the problem

Estimates of ability across the dataset



Graphical Test Using a Funnel Plot



Statistical Tests and Publication Bias

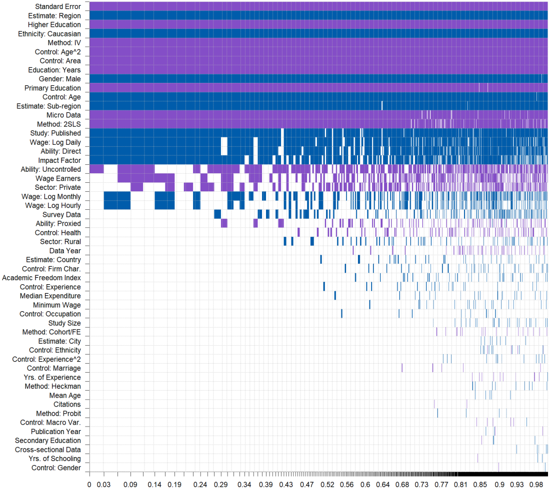
	OLS	FE	BE	RE	Study	Precision
Publication bias	0.832	0.746	0.752	0.747	1.169	0.262
(Standard error)	(0.097)	(0.060)	(0.244)	(0.058)	(0.121)	(0.425)
Effect beyond bias	6.408	6.517	6.741	6.708	6.294	6.540
(Constant)	(0.118)	(0.107)	(0.418)	(0.294)	(0.153)	(0.168)
	WAAP	**Top10**	**Stem**	**Hier**	**AK**	**Kink**
Publication bias				0.503	P = 2.764	0.262
				(0.168)	(0.107)	(0.39)
Effect beyond bias	6.9	6.439	7.2	6.801	6.548	6.54
	(0.092)	(0.548)	(1.186)	(0.266)	(0.091)	(0.054)
Observations	1,754	1,754	1,754	1,754	1,754	1,754

Individual Variables in Returns to Education

Over 30 variables split into six categories:

- Estimates and their descriptive statistics
- Estimate characteristics
- Data characteristics
- Spatial/structural variation
- Estimation method
- Publication characteristics

Model Inclusion in Bayesian Model Averaging



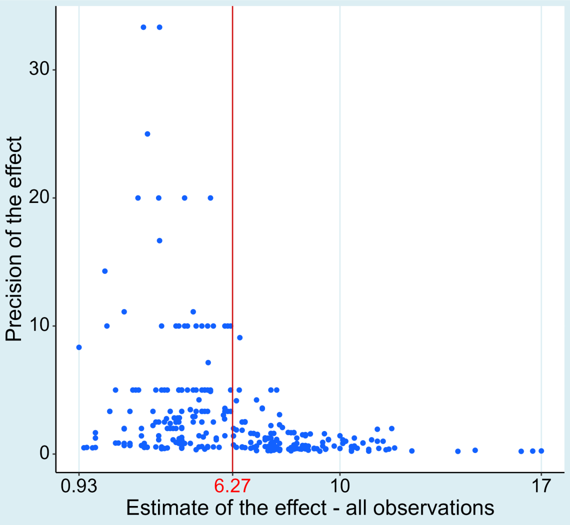
Economic Significance of Key Variables

	One SD change		Maximum change	
	Effect on Returns	% of BP	Effect on Returns	% of BP
Standard Error	0.642	9.82%	3.435	52.56%
Estimate: Sub-region	-0.428	-6.55%	-1.433	-21.92%
Estimate: Region	-0.612	-9.37%	-1.325	-20.27%
Education: Years	0.566	8.67%	1.175	17.98%
Wage: Log Daily	-0.405	-6.2%	-1.384	-21.18%
Micro Data	0.532	8.13%	1.391	21.29%
Primary Education	0.535	8.18%	3.540	54.16%
Higher Education	1.366	20.91%	5.521	84.48%
Gender: Male	-0.425	-6.5%	-1.215	-18.58%
Ethnicity: Caucasian	-0.608	-9.3%	-1.449	-22.18%
Method: 2SLS	0.433	6.62%	1.474	22.56%
Method: IV	0.824	12.61%	2.627	40.2%
Ability: Direct	-0.388	-5.94%	-1.138	-17.41%
Ability: Uncontrolled	0.271	4.15%	0.548	8.39%
Control: Age	-0.895	-13.69%	-1.883	-28.81%
Control: Age^2	1.315	20.12%	2.945	45.06%
Control: Area	0.878	13.44%	1.781	27.24%
Impact Factor	-0.296	-4.53%	-1.349	-20.64%
Study: Published	-0.445	-6.8%	-1.047	-16.01%

Making a twin dataset

- Only subjects with identical inherent ability - twins
- 16 twin studies with 293 observations
- Assumption: Differences in returns to education are due to differences in education

Twin Funnel Plot



Publication bias for twins

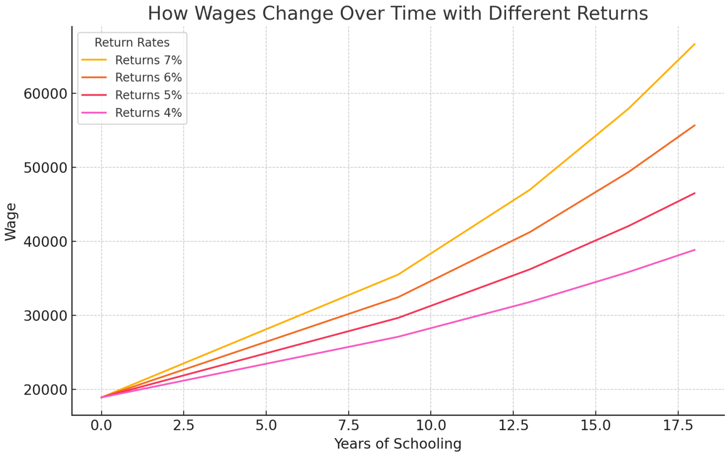
	OLS	FE	BE	RE	Study	Precision
Publication bias	1.347	0.602	2.133	0.840	0.947	2.897
(Standard error)	(0.138)	(0.162)	(0.505)	(0.154)	(0.177)	(0.442)
Effect beyond bias	4.735	5.574	4.106	5.55	4.754	3.907
(Constant)	(0.175)	(0.219)	(0.711)	(0.342)	(0.185)	(0.232)
	WAAP	**Top10**	**Stem**	**Hier**	**AK**	**Kink**
Publication bias				0.601	2.257	2.895
				(0.365)	(0.126)	(0.435)
Effect beyond bias	5.77	4.314	3.403	5.857	5.616	3.908
	(0.159)	(0.265)	(0.95)	(0.544)	(0.157)	(0.093)
Observations	293	293	293	293	293	293

Policy Implications

$$\log(w) = log(w_0) + \rho \times s$$

- w - wage
- w_0 - initial wage
- ρ - returns to schooling
- s - years in school

Ceteris Paribus



Conclusion

- An overall effect of returns to schooling drops roughly one percentage point (7% to 6%) after corrected for publication bias
- Ability matters, and controlling for it in the regression decreases the expected returns to schooling
- Nine variables have a significant positive influence on returns to schooling, while ten have a negative one
- The returns to schooling drop even further for twin studies with identical inherent ability (4% to 6%)

Thank you!

References

Mincer, Jacob A. "The human capital earnings function."
Schooling, experience, and earnings, pp. 83-96. NBER, 1974.

Blackburn, McKinley L., and David Neumark. "Omitted-ability bias and the increase in the return to schooling."
Journal of labor economics 11, no. 3 (1993): 521-544.

Schooling in Years vs. Levels

$$S_i = (1 + \beta_{i,higher} - \beta_{i,lower})^{\frac{1}{Y_{i,higher} - Y_{i,lower}}} - 1$$

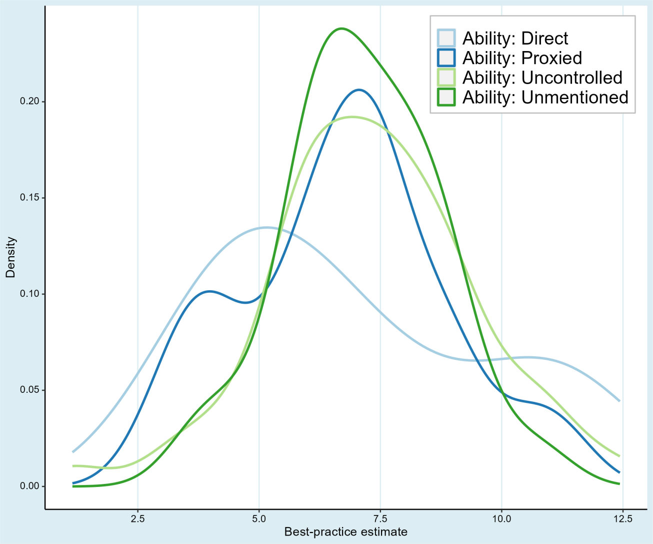
Results Using Some Recent Methods

Panel A: p-hacking tests by Elliott et al. (2022)				
	Non-increas.	**Monotonicity**		
Non-increas.	0.819	0.871		
Observations ($p \leq 0.1$)	1,610	1,610		
Observations	1,754	1,754		
Panel B: MAIVE estimator (Irsova et al., 2023)				
	Results			
MAIVE coefficient	5.736			
Standard Error	(0.460)			
F-test	12.491			
Observations	1,754			
Panel C: Robust Bayesian Model Averaging (Bartos et al., 2022)				
	Mean	**Median**	**0.025**	**0.975**
Coefficient	7.125	7.124	6.946	7.299
Standard Error	(3.505)	(3.504)	(3.371)	(3.645)
Observations	1,754	1,754	1,754	1,754

Best-Practice Estimate Across Literature

Study	Estimate	95% Confidence Interval	Studies
Author	6.536	(5.762; 7.310)	0
Query	7.529	(3.552; 11.506)	74
Snowballing	6.346	(2.530; 10.162)	41
All studies	7.109	(3.046; 11.17)	115

Aggregating BPE Results



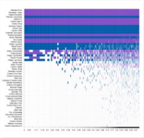
Meta-Analysis Automatization Script

- Meta-analysis, but automatized
- One script and a bit of parametrization
- Faster methods, all ran locally
- Caches and file handling
- All results calculated, formatted, and exported within minutes

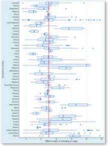
Project structure

```
.
├── data/
├── pckg/
├── scripts/
├── results/
│   ├── graphic/
│   ├── numeric/
│   └── main_results.txt
├── main_master_thesis_cala.R
├── script_runner_master_thesis_cala.R
├── source_master_thesis_cala.R
├── README.md
└── user_parameters.yaml
```

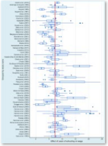
Graphic Results



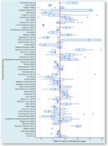
bma_UIP_uniform_results



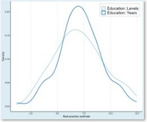
box_plot_country



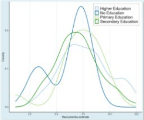
box_plot_study_name_1



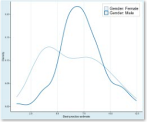
box_plot_study_name_2



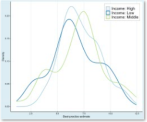
bpe_1



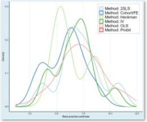
bpe_3



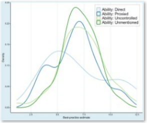
bpe_4



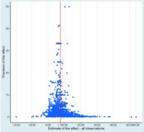
bpe_5



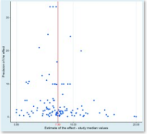
bpe_6



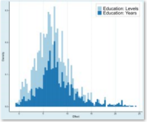
bpe_7



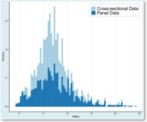
funnel



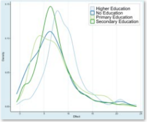
funnel_medians



prima_facie_1



prima_facie_2



prima_facie_3

Numeric Results

- bpe_econ_sig
- bpe_res_all_studies
- bpe_summary_stats
- effect_summary_stats
- exo_tests
- linear_tests
- ma
- ma_variables_description_table
- nonlinear_tests
- p_hacking_tests_caliper
- p_hacking_tests_elliott
- p_hacking_tests_maive
- robma_components
- robma_estimates
- variable_summary_stats

main_results.txt

```
[1] "Generating the prima facie graphs..."
[1] "Printing a box plot 1/2 for the factor: study_name"

[1] "Printing a box plot 2/2 for the factor: study_name"

[1] "Printing a box plot for the factor: country"

[1] "Results of the linear tests, clustered by study:"
                         OLS Fixed Effects Between Effects Random Effects Study weighted OLS Precision weighted OLS
Publication Bias    0.832***      0.746***        0.752***       0.747***           1.169***                  0.262
(Standard Error)     (0.097)        (0.06)         (0.244)        (0.058)            (0.121)                (0.425)
Effect Beyond Bias 6.408***      6.517***        6.741***       6.708***           6.294***                6.54***
(Constant)           (0.118)       (0.107)         (0.418)        (0.294)            (0.153)                (0.168)
Total observations      1754          1754            1754           1754               1754                   1754


[1] "Writing the linear tests results into ./results/numeric/linear_tests.csv"

[1] "Results of the non-linear tests, clustered by study:"
                       WAAP    Top10    Stem Hierarch Selection Endogenous Kink
Publication Bias                                0.503***  2.764***           0.262
(PB SE)                                          (0.168)   (0.107)          (0.39)
Effect Beyond Bias   6.9*** 6.439***  7.2*** 6.801***  6.548***         6.54***
(EBB SE)            (0.092)  (0.146) (1.186)  (0.266)   (0.091)         (0.054)
Total observations     1754     1754    1754     1754      1754            1754
Model observations     1469      176
```

It can do this...

- Variable summary statistics
- Effect summary statistics
- Prima Facie graphs
- Box plot
- Funnel plot
- T-statistic histogram
- Linear tests
  - OLS
  - Between Effects
  - Fixed Effects
  - Random Effects
  - Study-weighted OLS
  - Precision-weighted OLS

...and this...

- Non-linear tests
  - Weighted Average of Adequately Powered
  - Top10
  - Stem-based method
  - Hierarchial Bayes
  - Selection model
  - Endogenous Kink model
- Tests relaxing exogeneity
  - Instrumental Variable regression
  - p-uniform*
- P-hacking tests
  - Caliper tests
  - Elliott tests
  - MAIVE estimator

...and even this!

- Bayesian Model Averaging
- Frequentist Model Averaging
- Model Averaging variables description table
- Best-practice estimate
- Best-practice estimate: Graphs
- Best-practice estimate: Summary statistics
- Robust Bayesian Model Averaging

Available on GitHub

github.com/PetrCala/Diploma-Thesis