

REPLICATION OF “EDUCATIONAL
EXPANSION AND ITS HETEROGENEOUS
RETURNS FOR WAGE WORKERS”
BY
MICHAEL GEBEL AND FRIEDHELM PFEIFFER

Luisa Hammer and Marcelo Avila

22 Nov 2018

TODO: include outline of present.

THEORETICAL PART

INTRODUCTION

SUMMARY OF GEBEL & PFEIFFER (2010)

- basic idea: examine evolution of returns to education in West German labour market.
- Focus on change in returns to education over time as a consequence to education expansion in Germany.
- methodology:
 - Wooldridge's (2004) **conditional mean independence**
 - Garen's (1984) **control function** approach, that requires an *exclusion restriction*
 - as well as OLS
- data: SOEP 1984-2006

DATA AND VARIABLES

- Log of hourly wage
- Years of education (constructed from categorical variable)
- Age and age squared
- Gender
- Father's education
- Mother's education
- Father's occupation
- Rural or urban household
- Number of Siblings (as instrument)

TODO: more detailed table? (Comment) not necessary

BACKGROUND INFORMATION

- **increase in educational attainment** in the 1960s. From 1984 to 2006, average years of schooling increased:
 - woman: 11.3 \rightarrow 12.8
 - men: 11.9 \rightarrow 12.9
- **How can educational expansion affect the returns to education?**
- Standard theory: an increase of labor supply of high-skilled workers should decrease the returns to education
- High-educated workers with higher unobserved motivation / ability which positively affects wages
- More “less talented” accepted to higher education & thereby decreasing the average productivity levels of higher educated workers
 \rightarrow overall effect not clear
- unobserved characteristics leading to **selection bias**:
 - higher ability and motivation to stay longer in education

A FEW A PRIORI HYPOTHESES

(Comment) imo not super important, could be neglected because we have enough stuff tot alk about - but nice table!

Factors affecting RtE	↑ RTE	↓ RTE
Increase in female labour participation		✓
Birth cohort sizes (Baby boom)		✓
Wage determination processes (entrants)		✓
Skill-biased technological change	✓	

ECONOMETRIC APPROACH

EMPIRICAL FRAMEWORK (DERIVATION) I

The study is based on the **correlated random coefficient model** (Blundell / Dearden / Sianesi, 2005; Heckman / Vytlacil, 1998; Wooldridge, 2004).

$$\ln Y_i = a_i + b_i S_i$$

with $a_i = a'X_i + \varepsilon_{ai}$, and $b_i = b'X_i + \varepsilon_{bi}$

where $\ln Y_i$: log of wages and S_i years of schooling of individual i

- The model has, therefore, an **individual-specific intercept** a_i and **slope** b_i dependent on **observables** X_i and **unobservables** ε_{ai} and ε_{bi} .
- Does not assume that b_i and S_i are independent \rightarrow Individuals with higher expected benefits from education are more likely to remain longer in education $\rightarrow b_i$ may be correlated with S_i meaning positive self-selection.

EMPIRICAL FRAMEWORK (DERIVATION) II

- focus: estimate average partial effect (APE), which is the return per additional year of education for a randomly chosen individual (or averaged across the population)

$$E(\partial \ln Y / \partial S) = E(b_i) = \beta$$

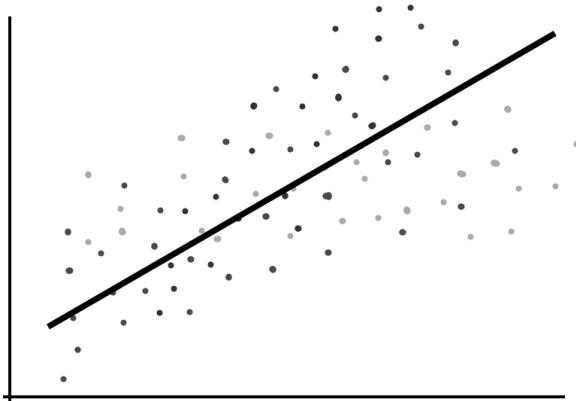
In case of homogenous returns to education the wage equation reduces to:

$$\ln Y_i = a'X_i + \bar{b}S_i + \varepsilon_{ai}$$

- Unobserved heterogeneity may only affect the **intercept** of the wage equation.

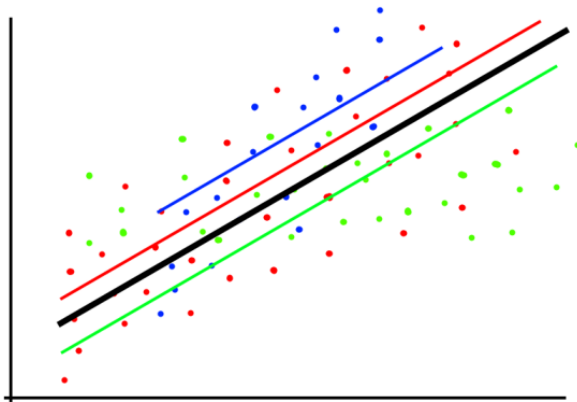
- still potential endogeneity if ε_{ai} correlates with S_i

EMPIRICAL FRAMEWORK (INTUITION) I



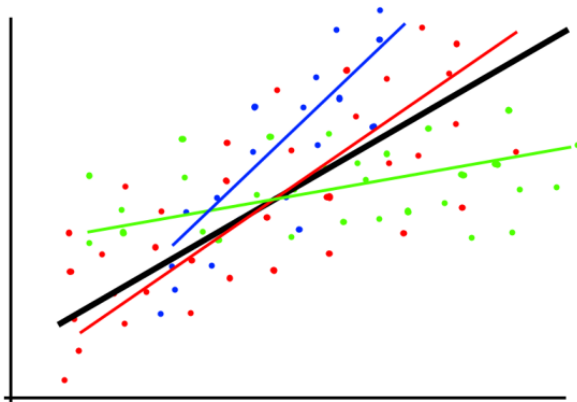
- Simple OLS

EMPIRICAL FRAMEWORK (INTUITION) II



- Multiple OLS with homogenous return to Educ

EMPIRICAL FRAMEWORK (INTUITION) III



- Correlated Random Coefficient Model

DISTINCTION TO CONVENTIONAL METHODS

- OLS
 - ability and “background” bias
- IV Methods
 - if education is correlated with **unobserved individual heterogeneity**, IV methods may fail to identify APE.
 - alternative: **Local Average Treatment Effect**.

CONDITIONAL MEAN INDEPENDENCE

According to Wooldridge (2004, pg. 7), **APE** is identified by:

$$E(\ln Y_i \mid a_i, b_i, S_i, X_i) = E(\ln Y_i \mid a_i, b_i, S_i) = a_i + b_i S_i \quad (\text{A.1})$$

$$E(S_i \mid a_i, b_i, X_i) = E(S_i \mid X_i) \text{ and } \text{Var}(S_i \mid a_i, b_i, X_i) = \text{Var}(S_i \mid X_i) \quad (\text{A.2})$$

TODO: add interpretation of assumptions

ESTIMATOR FOR β AND GLM

$$\hat{\beta} = \frac{1}{N} \sum_{i=1}^N \left((S_i - \hat{E}(S_i | X_i) \ln Y_i) / \hat{Var}(S_i | X_i) \right)$$

$$E(S_i | X_i) = e^{\gamma X_i} \quad \text{and} \quad Var(S_i | X_i) = \sigma^2 e^{\gamma X_i}$$

Where σ^2 can be consistently estimated by the mean of squared Pearson residuals and standard errors are bootstrapped.

CONTROL FUNCTION APPROACH I

- Based on proposition by Garen (1984).
- Similar to Heckman two-step estimator.
- Models schooling choice explicitly in first step
- CF approach can identify APE in heterogeneous returns while standard IV approach may not.

First stage: modelling of schooling choice

$$S_i = c'X_i + dZ_i + v_i \quad \text{with} \quad E(v_i \mid Z_i, X_i) = 0$$

where:

- X_i and Z_i influence the educational decision.
- v_i : Error term incorporating unobserved determinants of education choice.
- Z_i : Exclusion restriction (instrument).

CONTROL FUNCTION APPROACH II

- V_i , ε_{ai} and ε_{bi} are normally distributed with zero means and positive variances.
- possible correlation between error terms
- v_i is positive if an individual acquires higher education than expected conditional on observed characteristics

Second step: augmented wage equation

$$\ln Y_i = a_i + \beta S_i + \gamma_1 v_i + \gamma_2 V_i S_i + w_i$$

where:

- $\gamma_1 v_i$ and γ_2 are the **control functions**
 - $\gamma_1 = \text{cov}(\varepsilon_{ai}, v_i) / \text{var}(v_i)$
 - $\gamma_2 = \text{cov}(\varepsilon_{bi}, v_i) / \text{var}(v_i)$
- $E(w_i | X_i, S_i, v_i) = 0$ (as shown in Heckman / Robb, 1985)

CONTROL FUNCTION APPROACH III

Interpretation of the coefficients of the control functions - γ_1 measures the effect of those unobserved factors that led to over- or under-achievement in education on the wage - Thus, if γ_1 is positive, the unobserved factors affect schooling *and* wages positively - γ_2 describes how this effect changes with increasing levels of education - Positive coefficient would indicate that those with unexpected educational “over-achievement” tend to earn higher wages

TODO: intuition for CF approach

REPLICATION RESULTS



SET-UP

- We use the same sample: West Germans (not foreign-born or self-employed) between 25 & 60 years who work full-time
- We have less observations than Gebel & Pfeiffer (2010) per survey year after we delete all observations with missing values
- Yet, we extend the observation period until 2016
- Three estimation methods: OLS, CMI & CF
- We are not able to replicate the estimation results of Gebel & Pfeiffer (2010) exactly, yet the trend / shape is similar

RESULTS

- i'm not so sure how to add images / tables here but in the new do-file link <https://1drv.ms/u/s!Ap1Tm8513olthBjgylALS8Zp3A7G> you can just save the graph with all 3 approaches
- & then display on the other side the same graph from GP(2010, p.35)
- also: here is a table with the our & GP estimates for comparisons - your bootstrapped standard errors are already included

<https://1drv.ms/x/s!Ap1Tm8513olthBp5BPld0qO8h3Yj>

ESTIMATED RETURNS ON EDUCATION

- Estimates from OLS & CMI are similar, yet, CMI produces lower estimates which points to a positive self-selection bias
- Generally, CF estimates are much more volatile and less precise

Differences between replicated & original estimations - Our OLS estimates are on average larger than those of Gebel & Pfeiffer (2010) by 0.004 percentage points - Our CMI estimates are on average larger than those of Gebel & Pfeiffer (2010) by 0.002 percentage points (first years lower, than larger) - Our CF estimates are on average significantly larger by 0.032 percentage points, though the divergence gets smaller from 2000 onwards

CONTROL FUNCTION ESTIMATES

Instrumental variable in first stage - *number of siblings* is significant at the 0.1% level for all years - as expected, the number of siblings has a negative impact on the years of schooling (the estimates range between -0.13 & -0.23) - We would assume that the instrument does not directly affect the error term in the wage equation

Coefficients of the control functions - γ_1

EXPLANATIONS FOR DIVERGENCES BETWEEN REPLICATION AND GEBEL & PFEIFFER (2010)

- sample not the same
- ...

PRO'S & CON'S OF ESTIMATION METHODS

THE END :)