Bayesian vs. Frequentist Approach: A Comparative Study of Educational Returns to Income in Germany

VU UNIVERSITY AMSTERDAM

School of Business and Economics

Zoë Lyn Busche | Thesis Supervisor: Lennart Hoogerheide

Introduction

The importance of education goes well beyond the classroom and substantially influences many aspects of our lives. In this paper we focus on the effect of education on income in Germany. This relationship is not only important for understanding an individual's economic success but also reflects broader social dynamics such as access to opportunity and economic inequality.

Background information about the German school system (Wößmann et al., 2024):

- German school reproduces social inequalities in educational opportunities
- Parent's education strongly influences child's educational level independent of child's intelligence
- Children attending Gymnasium (highest secondary school): 26.7% with parents of lower education and income, 59.8% with parents of higher education and income

This paper analyses:

- Potential endogeneity of the variable education (Blackburn and Neumark, 1993)
- Validity of father's education as the instrumental variable
- Compare the Bayesian and frequentist approaches in analyzing the relationship between education and income

Endogeneity

Endogeneity occurs when the explanatory variable is correlated with the error term. If endogeneity is ignored, it can lead to biased results, and the true effect of the explanatory variable on the dependent variable is over- or underestimated. Possible causes of endogeneity are:

i. Omitted variables
ii. Measurement errors
iii. Reverse Causality

Criteria for Effective Instruments (Hoogerheide et al., 2012)

i. Exclusion restriction: Instrument should not directly affect the independent variable
ii. Strength of instrument: Instrument should have strong effect on endogenous explanatory variable
iii. Relevance: Instrument should have relevant effect

Methodology: Bayesian vs. Frequentist Approach

In the **Bayesian approach** parameters are no longer fixed but seen as random variables. A Bayesian statistician presumes that we cannot be sure what the true value of the unknown parameter is and incorporates prior assumptions or knowledge about the parameter into the model.

The Bayesian approach is based on Bayes' theorem (Bayes, 1763):

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

where θ is the unknow parameters and y the data.

Characteristics of Bayesian approach:

- Incorporates prior beliefs and uncertainty about unknown parameters
- Probability statements about unknown parameters
- Reliable results for finite samples
- Visible discrimination though choice of priori

The **frequentist approach** considers probabilities as the long-run frequencies of a process.

Characteristics of the frequentist approach:

- binary results
- perceived objectivity
- control over Type I errors through significance level
- reliable results for large sample size

Two-Stage Least Squares

The Two-Stage Least Square (2SLS) method is divided into two stages for j= $0, \dots, n$:

i. first-stage regression: $X_j = \pi_0 + \pi_1 Z_j + v_j$
ii. second-stage regression: $Y_j = \beta_0 + \beta_1 \widehat{X}_j + u_j$

with j the regressand, X_j the endogenous explanatory variable, Z_j the instrument and v_j and u_j the error terms.

Methodology: Gibbs Sampling

We use Gibbs sampling, a Markov Chain Monte Carlo method, to approximate the posterior density by using the conditional distribution (Geman and Geman, 1984). The sequence of samples forms a Markov chain whose stationary distribution is the posterior distribution. The process goes as follows:

i. The set of parameters θ is divided into m subsets $\theta_1, \dots, \theta_m$ and
$$\theta_{-s} = \{\theta_1, \dots, \theta_{s-1}, \theta_{s+1}, \dots, \theta_m\}$$
is defined as the set of $m-1$ subsets excluding the s-th subset.
ii. Feasible initial parameter values $\theta^0 = (\theta_1^0, \dots, \theta_m^0)$ are chosen.
iii. For draw $i = 1, \dots, n_{draws}$ and $s = 1, \dots, m$ we obtain $\theta_s^{(i)}$ from the conditional posterior density $p(\theta_s|\theta_{-s}^{(i-1)}, y)$ where
$$\theta_{-s}^{(i-1)} = \left\{\theta_1^{(i)}, \dots, \theta_{s-1}^{(i)}, \theta_{s+1}^{(i-1)}, \dots, \theta_m^{(i-1)}\right\}$$
denotes all subsets except for θ_s at their most recently simulated values.

We discard the first part of the Gibbs sequence by establishing a burn-in of 1,000 draws from 11,000 simulated draws.

Econometric Model

To estimate the effect of education on log(income), we specify the following with w_i the i-th control variable:

$$\log(income) = \alpha_1 + \beta\ education + \sum_{i=1}^{m} \delta_{1i} w_i + u_i$$

We use IV regression to address endogeneity. Implementing this idea, results in the following equation where z is the instrument:

$$education = \alpha_2 + \delta z + \sum_{i=1}^{m} \delta_{2i} w_i + u_2$$

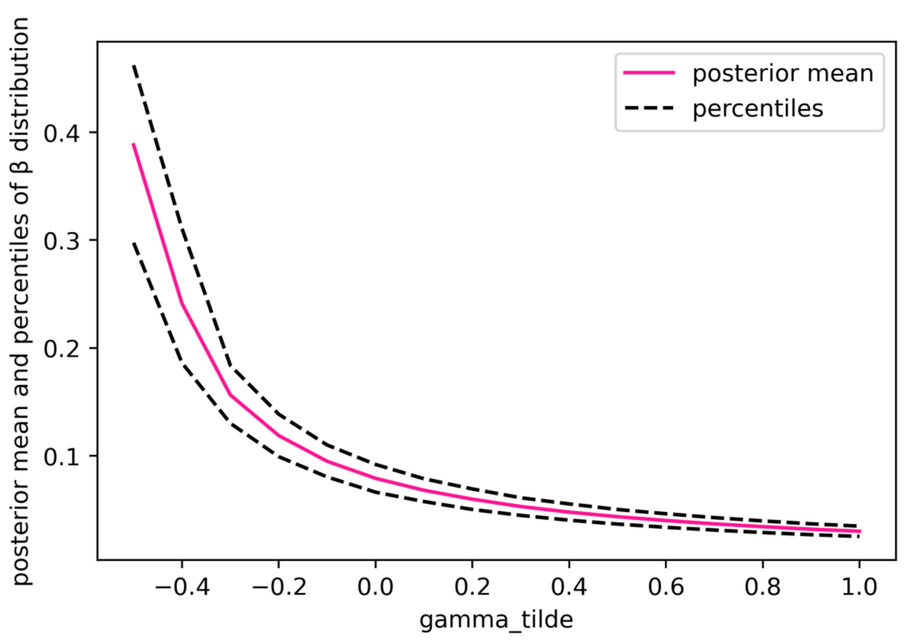
In the Bayesian approach we allow for a small direct effect (γ) of z on $income$. We redefine the model as:

$$\log(income) = \alpha_1 + \beta\ education + \gamma z + \sum_{i=1}^{m} \delta_{1i} w_i + u_i$$

We define the ration of the effect of the instrument the individual's own education on income as $\tilde{\gamma} = \gamma/\beta$.

Bayes, T. (1763). "An Essay towards Solving a Problem in the Doctrine of Chances". In: Philosophical Transactions 53, pp. 370–418.
Blackburn, M.L. and D. Neumark (1993). "Omitted-Ability Bias and the Increase in the Return to Schooling". In: Journal of Labor Economics 11.3, pp. 521–544.
Geman, S. and D. Geman (1984). "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". In: IEEE Transactions on Pattern Analysis and Machine Intelligence6.6, pp. 721–741.
Hoogerheide, L.F., J.H. Block, and R. Thurik (2012). "Are Education and Entrepreneurial Income Endogenous and do Family Background Variables make Sense as Instruments?" In: Entrepreneurship Research Journal 2.3, pp. 1–29.
Wößmann, L., F. Schoner, V. Freundl, and F. Pfaehler (2024). "Ungleiche Bildungschancen: Ein Blick in die Bundesländer". In: ifo Schnelldienst 77.5, pp. 49–62.

Results: Instrument Robustness Test



- $0 \leq \tilde{\gamma} < 1$: plausible because effect of own education expected to be larger than father's education
- $\tilde{\gamma} \to \infty$ then posterior mean $\to 0$
- $\tilde{\gamma} < 0$: implausible because effect of own education and father's education expected to have the same sign

→ Increase of posterior uncertainty not greater for relaxed exclusion restriction than for strict exclusion restriction.

$\tilde{\gamma}$	Mean	Std. Dev.	2.5%	97.5%
-0.5	0.3953	0.0526	0.3140	0.5145
0.0	0.0793	0.0065	0.0670	0.0922
0.2	0.0601	0.0049	0.0576	0.0697
50.0	0.0010	0.0004	0.0038	0.0011

Table 1 – Results of the posterior distribution of β for different values of $\tilde{\gamma}$.

Results: Comparing Methods

Method	Mean	Std. Dev.	2.5%	97.5%
OLS	0.0643	0.0020	0.0600	0.0690
2SLS	0.0793	0.0070	0.0660	0.0920
Bayesian IV	0.0793	0.0065	0.0670	0.0922

Table 2 – Estimates and statistics for OLS, 2SLS and Bayesian IV with $\tilde{\gamma} = 0$.

- Bayesian IV $(\tilde{\gamma} = 0)$ and 2SLS estimate implies an additional year of education results in an approximately 7.93% increase in income.
- Bayesian IV and 2SLS perform similarly well.
- Downward shift of the OLS estimate implies OLS underestimates true causal effect of education.

Conclusion

i. Bayesian and frequentist approach perform similarly well. However, the Bayesian approach offers a significant advantage: incorporating uncertainty about the exclusion restriction allowing for a robustness test of the instrument's validity.
ii. Family background variable is a valid instrument to estimate the return of education on income and criticism is unjustified.
iii. Data shows that income increases approximately by 7.93% for an additional year of education.

z.l.busche@student.vu.nl