

**SOEPpapers**  
on Multidisciplinary Panel Data Research

SOEP – The German Socio-Economic Panel Study at DIW Berlin

420-2011

**Smoking and Returns to Education –  
Empirical Evidence for Germany**

Julia Reilich

## **SOEPpapers on Multidisciplinary Panel Data Research** at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at  
**<http://www.diw.de/soeppapers>**

### **Editors:**

Joachim R. **Frick** (Empirical Economics)

Jürgen **Schupp** (Sociology, Vice Dean DIW Graduate Center)

Gert G. **Wagner** (Social Sciences)

Conchita **D'Ambrosio** (Public Economics)

Denis **Gerstorff** (Psychology, DIW Research Professor)

Elke **Holst** (Gender Studies)

Frauke **Kreuter** (Survey Methodology, DIW Research Professor)

Martin **Kroh** (Political Science and Survey Methodology)

Frieder R. **Lang** (Psychology, DIW Research Professor)

Henning **Lohmann** (Sociology, DIW Research Professor)

Jörg-Peter **Schräpler** (Survey Methodology, DIW Research Professor)

Thomas **Siedler** (Empirical Economics, DIW Graduate Center)

C. Katharina **Spieß** (Empirical Economics and Educational Science)

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)

DIW Berlin

Mohrenstrasse 58

10117 Berlin, Germany

Contact: Uta Rahmann | [soeppapers@diw.de](mailto:soeppapers@diw.de)

# Smoking and Returns to Education – Empirical Evidence for Germany

Julia Reilich\*

Looking at smoking-behavior it can be shown that there are differences concerning the time-preference-rate. Therefore this has an effect on the optimal schooling decision in the way that we assume a lower average human capital level for smokers. According to a higher time-preference-rate we suppose a higher return to education for smokers who go further on education. With our empirical findings we can confirm the presumptions. We use interactions-terms to regress the average rate of return with the instrumentvariable approach. Therefore we obtain that smokers have a significantly higher average return to education than non-smokers.

**Keywords:** Returns to education, Human Capital, Smoking-Effects

**JEL Classification:** J24, J31, I21

---

\*Universität Potsdam, Wirtschafts- and Sozialwissenschaftliche Fakultät, August-Bebel-Str. 89, 14482 Potsdam, Deutschland. E-Mail: reilich@uni-potsdam.de, Tel. +49-331-977-4636.

# 1 Introduction

Smokers differ from non-smokers with respect to their characteristics. Harmon et al. (2003) and Fersterer and Winter-Ebmer (2003) show a difference in the time-preference-rate. They examine that smokers have a higher time preference. Hence, by looking at a human capital investment, smokers will discount future income more than non-smokers. Therefore we expect a lower educational and income level on average for smokers. Levine (1997) verifies this presumption. He compares the income of both groups and find substantially differences in their wages. Khwaja et al. (2007) show in their analysis that smoking affects the duration of financial plannings, which demonstrates that for smokers the financial planning period is lower than for non-smokers. In human capital theory Mincer (1974) examines that to get the optimal individual schooling-decision the discounted life-time-earning has to be maximized. Therefore the discount factor plays an important role for this decision.

According to the differences in the time-preference for smokers we suppose 1) that they will achieve less education on average. In addition to this for smokers who accumulate human capital the potential return for the investment in education has to be higher than for non-smokers. Because of this reason we also suppose 2) a higher average return to education for smokers.

The remainder of the paper is organized as follows:

In section 2 we examine the standard-model by Jacob Mincer with special remarks to the coheny between the individual discount rate and the rate of return. In section 3 we estimate the average return for both groups. In the last part we deliver a summarizing conclusion.

# 2 Model

The theoretical basis for computing the returns to education are given by Mincer (1974), p. 9. He separates the productive life-time into two parts. First part is the time for accumulating universal education. During this period it is assumed that there is no income but the individuals have opportunity costs in terms of forgone earnings. Opportunity costs increases in schooling level. The second period is defined while workers accumulate firm-specific capital. During this time, individuals get a positive income according to their universal schooling level.

Fig. (1) shows the parted life-cycle.

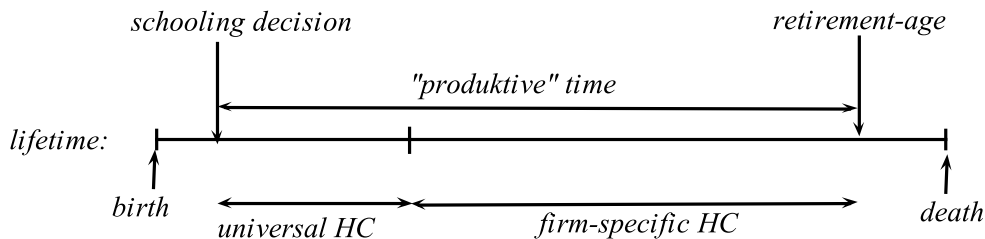


Figure 1: life-time-period

To determine the individual optimal schooling-level Mincer compares the present values from life-time-earnings for two alternative schooling-degrees.

For this simple model we solely assume that there are only differences in the annual

earnings because of different schooling levels. So we can establish the discounted life-time-earnings as follows

$$V_s = Y_s \sum_{t=s+1}^n \frac{1}{(1+r)^t}, \quad (1a)$$

where  $V_s$  is the discounted life-time-earning by a given universal schooling degree of  $s$  years.  $Y_s$  is the annual income, which is defined by the schooling degree,  $r$  represents the time-preference-rate, for which we assume to differ between smokers and non-smokers. For solvability Mincer assumes a constant working-period which starts after the universal schooling-time,  $s$ , with the first income after  $(s+1)$  years. If we assume that there are homogenous individuals by looking at ability-issues or family background we only get differences in the time preferences, which determine different schooling decisions.

For comparison the alternative choice of the schooling level is given by  $(s-d)$  years of schooling.

$$V_{s-d} = Y_{s-d} \sum_{t=s-d+1}^{n-d} \frac{1}{(1+r)^t}. \quad (1b)$$

However, individuals will decide to go further on education, if

$$V_{s-d} \leq V_s. \quad (2)$$

Solving the problem by assuming  $s$  equals  $d$  and by using equation (1a), (1b) and (2) we get the typical Mincer-equation for regressing the returns to education. So we get

$$\ln Y_s = \ln Y_0 + rs. \quad (3)$$

For our analysis the important fact of this derivation is that the time-preference-rate,  $r$ , from equation (1a) equals to the returns to education from equation (3). Therefore we should have a closer look to the relation between the return and the discount rate. For example for an individual with a high time-preference-rate there has to be a high potential return to education if the person will invest in human capital. In contrary if the person has a low time-preference-rate, the critical return on investment in education can be much lower. Therefore people with a lower discount rate will invest earlier in education than others. So, there is an inverse relation between the discount rate and the return to education. The individual will invest in education if the time-preference-rate is lower than the potential return on investment and will stop the investment by equalizing. In fig. (2) we show this relation more in detail.

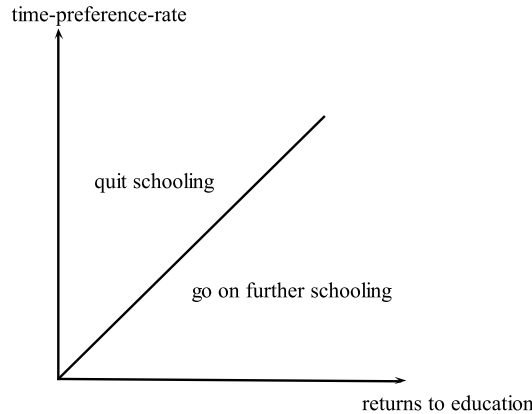


Figure 2: schooling decision

As said before, we can assume a difference in the average time-preference-rate with respect to smoking. Fersterer and Winter-Ebmer (2003) and others show that smokers have higher time-preference-rates than non-smokers. Based on that fact we assume that less smokers will go further on education, because the critical return is much higher. But those who invest in human capital will get a higher return on average.

### 3 Data and Results

We use data from the German-Socio-Economic-Panel collected by the German Institute of Economic Research<sup>1</sup>. Since 1984 they raise individual and household-data. By now, there are about 20.000 people in their sample. We take cross-section data from 2008 and reduce the sample to employees who work part- or full-time and who are not self-employed. To avoid a bias because of a later entry in the working period by high educated people we reduce the sample again to the age of 30 to 60. For our analysis there remain 4687 observations in our sample.

#### 3.1 Descriptive Statistics

In our sample there are 31% smokers. In tab. (1) we deliver descriptive statistics for the variables which we include in the models. We calculate the hourly wage from the given gross annual income and divide it by the numbers of working hours, fixed in the individual contract. As expected before we get a significant difference between smokers and non-smokers.

We can verify the hypothesis, that smokers earn less money than non-smokers. Similar results we get for the variable of the years of education. This is a generated variable by DIW. The codification is given in the documentation of the institute. However, the important information is, that for each educational-level which is achieved there is only one possible number of years of education. There is no bias, for example because of people who need longer time for their studies than the average. Therefore as we supposed we get a much higher average schooling-level for non-smokers. Experience is the potential experience following Mincer. We calculate the age minus the years of schooling minus 6 and minus the number of years with the experience of being unemployed. Female and married are both dummy variables with a value of 1, if so and zero otherwise. However, in addition we implement an interaction-term for married females which get the value 1 if a woman is married. The three variables big<sup>2</sup> firm, commuting and living in a rural<sup>3</sup> area are all dummy variables as well. Comparing the means from tab. (1) we can see huge differences between smokers and non-smokers. By using a t-test for two samples on each variable we evaluate on the 1%-level a significant difference between the means for every variable. For the numbers of brothers and sisters we count significant higher values for smokers. Same apply for number of migrants in this group. The rest of the control variables show higher value for non-smokers. Similiarly behave the number of women in the sample or the number of employees in a big firm.

---

<sup>1</sup>Shortly: DIW.

<sup>2</sup>Big firm is defined as a company with more than 200 employees.

<sup>3</sup>A rural area is defined as that the individual residence is at least 25km away from the next core of a city.

variable	mean		95% conf. interval	
	smokers	non-smokers	smokers	non-smokers
hourly wage ( $Y_i$ )	16.264 (0.229)	18.303 (0.169)	15.816- 16.712	17.971- 18.635
years of education ( $S_i$ )	11.938 (0.061)	13.138 (0.050)	11.818- 12.058	13.040- 13.237
experience ( $exp_i$ )	25.765 (0.218)	26.062 (0.155)	25.338- 26.193	25.757- 26.367
female ( $fem_i$ )	1.411 (0.013)	1.500 (0.009)	1.386- 1.436	1.483- 1.518
married ( $marr_i$ )	0.615 (0.013)	0.734 (0.008)	0.590- 0.640	0.719- 0.749
female married ( $IAT_{fem-marr_i}$ )	0.226 (0.011)	0.358 (0.008)	0.205- 0.248	0.341- 0.374
big firm ( $D_{bigfirm_i}$ )	0.472 (0.013)	0.510 (0.009)	0.447- 0.498	0.493- 0.527
rural ( $D_{rural_i}$ )	0.376 (0.013)	0.392 (0.009)	0.352- 0.401	0.375- 0.408
commute ( $D_{commute_i}$ )	0.556 (0.013)	0.564 (0.009)	0.531- 0.582	0.547- 0.581
fath. educ.: middle ( $D_{ME_i}$ )	0.264 (0.011)	0.288 (0.008)	0.241- 0.286	0.272- 0.304
fath. educ.: high ( $D_{HE}$ )	0.037 (0.005)	0.070 (0.005)	0.028- 0.047	0.061- 0.079
# of books ( $D N Books_i$ )	0.242 (0.011)	0.345 (0.008)	0.220- 0.264	0.329- 0.361
# of sisters ( $D N Sisti_i$ )	0.961 (0.030)	0.826 (0.018)	0.902- 1.020	0.791- 0.860
# of brothers ( $D N Broth_i$ )	0.985 (0.031)	0.833 (0.018)	0.923- 1.047	0.798- 0.868
migration ( $D_{Migr_i}$ )	1.077 (0.007)	1.045 (0.004)	1.063- 1.090	1.038- 1.052
observations	1472	3215		

Table 1: descriptive statistics; Source: DIW, own calculations, standard deviation in parenthesis.

### 3.2 Estimation

We define a smoker as an individual who smoked actively in 2008. There might be problems with changing the smoking-behaviour, but we assume, that individuals who smoke and are aged between 30 to 60 have the typical characteristics. To run the regression we start with an OLS-estimation by using the Mincer-equation. Furthermore we imply some control variables to reduce the heterogeneity of the individuals. Our first estimation-model is shown in the following equation

$$\begin{aligned} \ln Y_i = & \beta_0 + \beta_1 S_i + \beta_2 exp_i + \beta_3 exp_i^2 + \beta_4 fem_i + \beta_5 marr_i + \\ & \beta_6 IAT_{fem-marr_i} + \beta_7 D_{bigfirm_i} + \beta_8 D_{rural_i} + \beta_9 D_{commute_i} + \varepsilon_i. \end{aligned} \quad (1)$$

The results of the estimation are shown in tab. (2). According to Card (2001) by using OLS we have to test whether the schooling variable is endogenous because of the ability-bias. If so, both the schooling and earnings-variable is influenced by ability and motivation. This influence is hard to measure, so that we cannot introduce a variable in our model. If schooling is endogenous we have to use the IV-method to get an unbiased estimator. Therefore in the first step we run a regression on the schooling-variable. For this we imply instruments in the Reduced-Form-Regression. However, these are sufficient, if they are correlated with schooling but uncorrelated with income and ability. Obviously it is hard to find good instruments for this model. In our specification we use the following model:

$$\begin{aligned} S_i = & \gamma_0 + \gamma_1 D_{ME_i} + \gamma_2 D_{HE_i} + \gamma_3 N Books_i + \gamma_4 N Sisti_i \\ & + \gamma_5 N Broth_i + \gamma_6 D_{Migr_i} + \sum_{j=7}^{14} \gamma_j X_{ij} + \eta_i. \end{aligned} \quad (2)$$

Therefore the variable-descriptions are given in tab. (1).  $X_{ij}$  are the control variables which are also implemented in the wage equation. With the Reduced-Form-Regression we regress a potential number of years of schooling without ability-issues. In the second step we imply the estimated schooling levels from the first step and regress the wage function again.

Comparing the returns to education we get for both regressions a slightly higher return for non-smokers than for smokers. By using OLS we get a return about 7.9 %<sup>4</sup> for smokers and 8.7 % for non-smokers. Using IV we get 9.4 % for smokers and 10.7 % for non-smokers. Because of the separation of the sample it is not possible to test the results for equality of the coefficients. Because of this problems we cannot be satisfied with our results because the coefficients are nearly equal. As a proxy for similarity we compare the 95 % confidence intervals. Therefore we cannot establish that they are significantly unequal for both OLS and the IV-regression. From both models, we cannot verify our hypothesis, that smokers have different or even higher returns than non-smokers. By comparing the returns between both regressions we normally would expect that because of the ability-bias in the OLS-regression the return to education will be overestimated. In contrary to this we find, like in most other researches, that OLS-estimator is lower than the return regressed by IV. The reason for that are measurement errors in the schooling variable. Griliches (1977) discusses this problem and shows that this error will lead to an underestimation of the return to schooling by using OLS. Using IV or 3SLS does not give different results. So we reduce this output and show only IV-results. The necessity of IV can be tested by using the Durbin-Wu-Hausman-Test. Therefore we clearly have to reject the hypothesis that schooling is exogenous. As a next step we have to test the validity of our instruments. Therefore we use the approach achieved from Bound et al. (1995). First we have a look at the F-Statistic of the Reduced-Form-Regression. Therefore we can reject the hypothesis that the instruments have no impact on schooling. The next step is to look at the F-Test from another regression. By using IV we try to eliminate the relation between schooling and the error term from the wage equation. Therefore we run a regression by estimating

---

<sup>4</sup>To count the percentage return we calculate:  $(e^{\beta_1} - 1) \cdot 100$ .

the instruments on this error term. Therefore the instruments are valid, if the hypothesis of an impact of the instruments on the error term is excluded and so if we cannot reject the hypothesis based on the F-statistics.

variables	OLS, model 1		IV, model 2		IV, model 3
	smokers	non-smokers	smokers	non-smokers	
years of education	0.076*** (0.005)	0.083*** (0.003)	0.090*** (0.008)	0.102*** (0.004)	–
Schooling smokers	–	–	–	–	0.135*** (0.044)
Schooling non-smokers	–	–	–	–	0.088*** (0.004)
experience	0.035*** (0.007)	0.029*** (0.004)	0.037*** (0.007)	0.031*** (0.004)	0.032*** (0.004)
experience <sup>2</sup>	-0.000*** (0.000)	-0.000*** (0.000)	-0.001*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)
female	-0.111*** (0.034)	-0.141*** (0.026)	-0.113*** (0.034)	-0.148*** (0.027)	-0.130*** (0.021)
married	0.162*** (0.030)	0.080*** (0.023)	0.162*** (0.030)	0.072*** (0.023)	0.110*** (0.019)
female married	-0.188*** (0.044)	-0.119*** (0.031)	-0.186*** (0.044)	-0.111*** (0.031)	-0.142*** (0.026)
big firm	0.235*** (0.021)	0.214*** (0.014)	0.230*** (0.022)	0.210*** (0.014)	0.214*** (0.012)
rural	-0.100*** (0.022)	-0.032** (0.141)	-0.100*** (0.022)	-0.024* (0.014)	-0.049*** (0.012)
commute	0.078*** (0.022)	0.055*** (0.014)	0.077*** (0.022)	0.056*** (0.014)	0.063*** (0.012)
smokers					0.582 (0.706)
const	1.218*** (0.118)	1.309*** (0.077)	1.014*** (0.153)	1.002*** (0.096)	-0.021 (1.233)
<i>adj. R</i> <sup>2</sup>	0.31	0.36	0.30	0.35	0.32
F-Test	79	224	56	144	164
<i>First stage</i>					
<i>adj. R</i> <sup>2</sup>	–	–	0.35	0.44	0.95/ 0.91
F-statistics	–	–	57	178	6148/ 3237
F-statistics (error term)	–	–	2.18	1.07	0.94
Hausman (p-value)	–	–	0.000	0.000	0.000

Table 2: Estimation-results, source: DIW, own calculations, standard deviation in parenthesis, sign. level: \* 10 %, \*\* 5 % and \*\*\* 1 %.

For our regression model, we have to reject the hypothesis for smokers, but not for non-smokers. That shows a validity of the instrument solely for bigger samples. Unfortunately we cannot show that we use valid instruments for the separated regression for smokers. Because of this, we introduce a new model to identify the returns to education.

The aim of the new model is 1) to get the possibility to test the difference of the coefficients and 2) to verify valid instruments. Therefore we use a third model:

$$\begin{aligned} \ln Y_i = & \beta_0 + \beta_1 IAT_{Smoker-S_i} + \beta_2 IAT_{non-Smoker-S_i} + \beta_3 exp_i & (3) \\ & + \beta_4 exp_i^2 + \beta_5 fem_i + \beta_6 marr_i + \beta_7 IAT_{fem-marr_i} + \beta_8 D_{bigfirm_i} \\ & + \beta_9 D_{rural_i} + \beta_{10} D_{shuttle} + \beta_{11} D_{smoker} + \varepsilon_i. \end{aligned}$$

Here we imply two interaction terms, where the first term shows the years of schooling just for smokers, the second term just for non-smokers. In addition to this we introduce a dummy variable which controls for the pure income-effect of smoking. This introduces a multicollinearity-problem which has to be accepted. Again, we run this regression by using IV, because of an endogeneity problem for the interaction terms which we can establish. In contrast to the first models we get higher returns to education for smokers, as we expected before. So, smokers get a return from about 14.5 %, while non-smokers have on average a return from about 9.2 %. Because of not-separating the sample now we can test the difference of the coefficients by Wald-Test. Therefore we significantly can reject the hypothesis that the coefficients are equal. Because of the given multicollinearity we cannot be sure about the absolute value of the coefficients. But, the established difference is that big, that we can assume that the direction of the relation is true. Testing the validity of the instruments we get a rejection of the hypothesis, that the instruments cannot explain the interaction terms of schooling. For the Reduced-Form-Regression we find a very high adjusted  $R^2$  because of the big correlation between the smoker-dummy and the interaction-terms. Furthermore we run a regression of the instruments on the residuals of the wage equation. Therefore now there is no way to reject the hypothesis for no correlation between the error term and the instruments. For this reason we can conclude using valid instruments for the Reduced-Form-Regression with interaction terms and we can correct for the ability bias problem. Within our third model we can solve the problems from model 2 and in addition to this we find evidence for our assumed relations between the returns to education for smokers and non-smokers.

## 4 Conclusions

First of all we supposed differences in the schooling level on average. By looking at the descriptive statistics we can verify this assumption. As a second hypothesis we tried to verify that a) there are differences between the returns to education for smokers and non-smokers and if so b) that smokers do have higher returns. The reason for that assumption is that smokers have a higher time-preference-rate, so that they need to get higher returns to education to go further on accumulating human capital. We expected, that less smoking-people will go further on education, but those who accumulate human capital should get higher returns. Using the typical Mincerian-approach separately for both groups we cannot estimate a difference between the average return to education for smokers and non-smokers. Because of the separation of the sample, there is no way to test the coefficients, but by comparing the 95 % confidence intervals we get very similar intervals for both coefficients. In addition to this we are concerned about the ability bias and run a regression with the

IV-method. Therefore the coefficients rise for both groups but the differences are similar. In addition to this we cannot establish the validity of the instruments for the smokers. However, as another estimation we use a model, for that we can test the difference of the coefficients. According to this, the coefficient for the return to education for smokers is much higher than for non-smokers. Now we establish the use of valid instruments and can count a significant difference between smokers and non-smokers. Using our data, we can accept our hypothesis, that the return to education differs between smokers and non-smokers. In addition to this, we establish that a smoker only goes further on education if his expected return to education is higher than the critical value needs to be for a non-smoker.

## References

- [1] Bound, J., Jaeger, D.A. and Baker, R.M. (1995). 'Problems with Instrumental Variable Estimation when the Correlation between the instruments and the Endogenous Explanatory Variable is weak.' *American Statistical Association*, Bd. 90, N. 430, S. 443-450.
- [2] Card, D. (2001). 'Estimating the Return to Schooling: Progress on some Persistent Econometric Problems.' *Econometrica*, Bd. 69, N. 5, S. 1127-1160.
- [3] Fersterer, J. and Winter-Ebmer, R. (2003). 'Smoking, discount rates, and returns to education.' *Economics of Education Review*, Bd. 22, S. 561-566.
- [4] Griliches, Z. (1977). 'Estimating the returns to schooling: some econometric problems.' *Econometrica*, Bd. 25, Nr. 1, S. 1-21.
- [5] Harmon, C., Oosterbeek, H. and Walker, I. (2003). 'The Returns to Education: Microeconomics.' *Journal of Economic Surveys*, Bd. 17, 2, S. 115.
- [6] Khwaja, A., Silverman, D. and Sloan, F. (2007). 'Time Preferences, Time Discounting, and Smoking Decisions.' *Journal of Health Economics*, Bd. 26, 5, S. 927-949.
- [7] Levine, P.B. (1997). 'More Bad News for Smokers? The Effects of Cigarette Smoking on Wages.' *Industrial and Labor Relations Review*, Bd. 50, 3, S. 493-509.
- [8] Mincer, J. (1974). *Schooling, Experience and Earnings*. National Bureau of Economic Research.
- [9] Oosterbeek, H. and Ophem, H.v. (2000). 'Schooling choices: Preferences, discount rates, and rates of return.' *Empirical Economics*, Bd. 25, S. 15-34.
- [10] Suter, S. (2005). 'Bildung oder Begabung? Eine ökonometrische Untersuchung zu den Ursachen individueller Lohnunterschiede.' *Nationales Forschungsprogramm Bildung und Beschäftigung*, Bd. 43.