RUHR
ECONOMIC PAPERS

Jan Kleibrink

Causal Effects of Educational Mismatch in the Labor Market

RUHR UNIVERSITÄT BOCHUM RUB

#421

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #421

Responsible Editor: Thomas K. Bauer

ISSN 1864-4872 (online) – ISBN 978-3-86788-477-8

Ruhr Economic Papers #421

Jan Kleibrink

Causal Effects of Educational Mismatch in the Labor Market

RUHR UNIVERSITÄT BOCHUM

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788477
ISSN 1864-4872 (online)
ISBN 978-3-86788-477-8

Jan Kleibrink[1]

Causal Effects of Educational Mismatch in the Labor Market

Abstract

This paper analyzes the effect of educational mismatch on wages in Germany, using data from the German Socio-Economic Panel. Educational mismatch has been discussed extensively, mostly by applying OLS wage regressions which are prone to an unobserved heterogeneity bias. This problem is approached by using FE and IV models. As a stability check, the regressions are rerun using data from the International Adult Literacy Survey, allowing for an explicit control of skills as proxy of abilities. Results show that unobserved heterogeneity does not explain the wage differences between actual years of education and years of required education. This rejects the hypothesis that mismatched workers compensate for heterogeneity in innate abilities. The results suggest a structural problem in the German educational system as skill demand and supply are not in long-term equilibrium.

JEL Classification: I14, I21, J31

Keywords: Wages; educational mismatch

July 2013

1 Ruhr Universität Bochum. – The author thanks Prof. John P. Haisken-DeNew and Dr. Mathias Sinning for highly useful comments and suggestions. – All correspondence to Jan Kleibrink, RUB, RUB Research School, Room GC 2/148, Universitätsstr. 150, 44801 Bochum, Germany, E-Mail: jan.kleibrink@rub.de.

1 Introduction

In Germany, as well as in many other Western societies, the mean educational attainment of the population has constantly increased over the past decades. At the same time, the phenomenon of overeducation, a situation in which individuals have more education than they need for their jobs, has become increasingly relevant.

Overeducation is an extensively discussed topic in the economic literature (for an overview, see Hartog, 2000; Sicherman, 1991). The consequences of overeducation were analyzed in different settings including wage (Sicherman, 1991; Groeneveld and Hartog, 2004) and job satisfaction (Battu, Belfield, and Sloane, 1999) regressions. Although the topic has been analyzed in a vast amount of studies and over a long period of time, several conceptual issues still remain unsolved. While the operationalization of over- and undereducation has received substantial attention (see e.g. Verdugo and Verdugo, 1989; Rubb, 2003), the issue of causality has been addressed only recently (see e.g. Bauer, 2002; Korpi and Tahlin, 2009). However, understanding the causality in the relationship between educational match and wages is a necessary prerequisite for a complete understanding of the underlying mechanisms. This study addresses the problem of unobserved heterogeneity, which is still a common concern in this field of literature. As the explanation of the effects of mismatches is still inconclusive in the literature, this paper directly contributes to this discussion.

The standard way of operationalizing the educational mismatch framework compares the educational attainment of an individual to the educational attainment of workers in the same occupation (classical OMU framework). This framework was introduced by Duncan and Hoffman (1981) and has become widely accepted in the mismatch literature (e.g. Rubb, 2003). The OMU model explicitly differentiates between all three possible educational matches: overeducation (O), educational match (M) and undereducation (U), all of these states measured in years of education. OLS results, which are remarkably stable over nearly all studies in this field, show that overeducation has a positive significant effect on wages. Hence, overeducated individuals earn more than those less educated who are in the same occupation. However, the positive overeducation effect is significantly smaller than the effect of matched years of education. Hence, overeducated individuals are disadvantaged in terms of wages compared to others with the same education who are in a better match. Undereducation is either significantly negative or insignificant (a summary is found in Hartog, 2000). Even less consensus exists regarding the role of unobserved heterogeneity. Using overeducation in classical OLS wage regressions relies on the assumption that equally educated individuals have the same innate ability and thereby productivity (given other controls). However, this is a very strong assumption and has been criticized in mismatch studies lately (Bauer, 2002; Korpi and Tahlin, 2009). There are three ways of solving this problem: Fixed effects regressions (Bauer, 2002), the use of instruments (Korpi and Tahlin, 2009) and the direct inclusion of ability controls (Korpi and Tahlin, 2009). Each of these approaches has strengths and weaknesses (a discussion of the these approaches can be found in section 3). Therefore, concentrating on one of them alone involves the danger of interpreting results that are heavily driven by the model assumptions. I compare the results of three different methods offering a broad and robust base for interpretation.

The controversy regarding the role of unobserved heterogeneity leads to a problem in the interpretation of the underlying causes of the findings. A prominent approach is the human capital hypothesis (Hartog and Oosterbeek, 1988), which was tested for the German labor market by Bauer (2002). The idea is that mismatched individuals are in a "bad match" regarding their formal education because they compensate for disadvantages in their innate ability. When controlling for unobserved heterogeneity in the estimations, the coefficients of all three states should come close to each other to prove this hypothesis. Bauer (2002) finds that the compensation hypothesis cannot be completely rejected for Germany.

In the first part of the analysis, I replicate the OLS wage regressions with my sample. These standard OLS regressions replicate the findings that are established in the literature. However, there might be bias due to unobserved heterogeneity. Fixed effects results hint at the validity of the human capital compensation hypothesis as also found by Bauer (2002) using the same data set. However, IV regressions show that unobserved heterogeneity does not explain wage differences between years of required education and years of over-/undereducation unambiguously. Using an IV approach on SOEP data and including formerly unobserved skills as proxies of innate ability by using data from the IALS shows that differences between educational match and mismatch even become larger and only the years of required education matter, while years of overeducation do not have any significant wage effect. The finding that overeducation does not pay off in terms of wages neglects the explanation that mismatched workers compensate for skill shortages. These results show a structural problem in the German educational system because many individuals stay in the educational sector longer than they need to perform their future tasks. This imposes high costs for individuals, as they face foregone wages, as well as for the society because education is publicly funded to a large extend in Germany.

This analysis contributes to the economic literature in several dimensions: (1) By applying three different strategies of dealing with unobserved heterogeneity, it offers the most comprehensive study on the causal effects of overeducation in Germany; (2) the results contrast former results for the German labor market, offering a new line of argumentation; (3) this is of huge political relevance at it identifies a large-scale problem in the allocation of individuals in the educational sector and in the labor market.

The paper is organized as follows: Section 2 places this study in the existing literature, section 3 explains the empirical methods applied. Section 4 introduces the dataset used and explains the key variables. In section 5, the empirical results are explained and discussed, a conclusion can be found in section 6.

2 Literature

Educational mismatch has received substantial attention in the economic literature, mainly from an empirical point of view. Although already analyzed from various different angles, there are still gaps to be filled. Neither has the problem of unobserved heterogeneity been solved conclusively, nor has an explanation of the empirical findings in the literature found broad acceptance.

Duncan and Hoffman (1981) started the modern empirical mismatch literature by introducing a framework in which individual education consists of three parts: Education required for a job, overeducation and undereducation. This was the starting point of the OMU theory. A well-matched worker has exactly the years of education required in his job (M); overeducated workers attained additional years of education which are not needed for their current working life (O); undereducated workers received less education than required to do their jobs (U).

By decomposing obtained education into these three parts, it is possible to analyze if education generally pays off in the labor market or if it matters whether this education is used productively in an educational match. The OMU framework has become the standard approach in the overeducation literature and is, for example, used by Sicherman (1991); Rumberger (1987); Alba-Ramirez (1993); Bauer (2002); Hartog and Oosterbeek (1988) and Korpi and Tahlin (2009).

While these studies differ regarding their definition of overeducation, the datasets applied and the countries and time periods studied, they all share a common finding, which has become a stylized fact in the mismatch literature. Required education is positively associated with wages, the same is true for overeducation but the overeducation coefficient is significantly smaller. Undereducation is normally negative and significant but this finding is not as robust as the other two (for an extensive meta-analysis of the literature, see Hartog, 2000; Rubb, 2003).

The classical findings of mismatch studies are mainly based on OLS wage regressions. However, more recent papers have started using different regression techniques to tackle a possible problem of earlier studies: unobserved heterogeneity. Studies by Bauer (2002) and Korpi and Tahlin (2009) assume that unobserved heterogeneity biases OLS results, and discuss this in the context of the human capital theory. This theory assumes that educational mismatch is not a result of a structural mismatch of skill demand and supply in the labor market but mismatched workers compensate for ability not captured by the educational attainment.

Overeducated workers lack ability and compensate for this by getting more education than they actually need to perform their job. The opposite is true for undereducated workers. As they have a higher innate ability than others, they can get better jobs without having the required educational attainment.

According to Bauer (2002), this assumption is proven right when the coefficients of the three components of education become more equal when controlling for unobserved heterogeneity. Using data from the German SOEP, he finds that this hypothesis cannot be generally rejected for Germany as the coefficients become similar in size using panel models. Korpi and Tahlin (2009) do not only apply these panel models but also an IV model and a direct inclusion of skills. In their comprehensive approach, they reject the human capital theory but they state that in their instrumental approach, they are concerned with a weak instrument problem.

Using the SOEP, which runs over a longer period of time, I can get more robust results from panel models and thereby replicate the results found by Bauer (2002). I then extend this study by following the strategy of Korpi and Tahlin (2009), applying an IV approach and including skill measures directly. I use different instrument variables to avoid the abovementioned weak instrument problem. In a further step, I apply data from the International Adult Literacy Survey (IALS), a dataset explicitly designed to model skills. By using this dataset, I can directly include these skills to proxy ability in the OMU regressions and thereby avoid unobserved heterogeneity. Using this strategy, I contrast the results by Bauer (2002) and offer a different explanation for the existence of overeducation in the German labor market.

3 Estimation Method

Different estimation methods are applied in this analysis to identify causal effects. The starting point for wage analyses are Mincer wage regressions (Mincer, 1970). This approach is adjusted by including the variables for required education, overeducation and undereducation. All of these variables are measured in years. This gives the wage regression for the classical OMU framework (e.g. Hartog, 2000):

$$ln(w_{it}) = F(x_{it}\beta, O_{it}\gamma, M_{it}\zeta, U_{it}\eta) = \beta_0 + x_{it}\beta + O_{it}\gamma + M_{it}\zeta + U_{it}\eta + \varepsilon_{it} \quad (1)$$

with the logarithm of hourly wage $ln(w_{it})$ as the dependent variable. The main explanatory variables are educational match (M), overeducation (O) and undereducation (U) in years. The matrix x_{it} includes further controls such as a polynomial in age, a dummy for the birth cohort, tenure, the number of children and marital status, nights spent in hospital as control for individual health, a dummy for fulltime employment as well as industry and year dummies. ε_{it} is the error term.

The question arising here is how to operationalize overeducation. Basically, there are three different options: (1) An objective approach relying on an expert valuation; (2) a subjective approach relying on workers' self assessment; (3) an empirical approach. The first option is mainly used in US studies (e.g. Rumberger, 1987; McGoldrick and Robst, 1996). The Dictionary of Occupational Titles (DOT) for the US labor market provides the necessary information, such as information on occupations and the required education for these occupations. This assessment is made by labor market experts. An analogous does not exist for Germany. However, its unavailability is not the only reason not to use the "expert" method. Kiker, Santos, and de Oliveira (1997) point out that this measure can only work if updated regularly to account for technological change as well as hiring standards. If not updated regularly, this measure increasingly tends to misclassify respondents over time.

The subjective approach (e.g. Sicherman, 1991; Sloane, Battu, and Seaman, 1999) has the advantage over the expert valuation of being updated automatically with each wave of a panel dataset. Respondents are asked for the qualification necessary to do their job and the answer is compared to the actually attained education to evaluate whether a person is in an educational match or not. However, this definition gives rise to other problems. The main criticism of this approach is that it remains unclear which benchmark is used by respondents. They could either use the qualification necessary to actually perform the job, or they could answer according to hiring standards (Bauer, 2002). Whichever answer they give, it requires respondents to know about the standards in their occupational field in detail.

This study applies the empirical method of measuring overeducation, which was applied by e.g. Verdugo and Verdugo (1989); Kiker, Santos, and de Oliveira (1997) and Bauer (2002). Verdugo and Verdugo (1989) introduced this measure by using the mean value for each occupation. Overeducated workers are those individuals whose education exceeds the mean value plus one standard deviation, undereducated are those whose education lies below the mean value minus one standard deviation. This approach offers several advantages over the ones discussed before: Firstly, it is naturally updated regularly, similar to the subjective approach. Secondly, it does not have the weakness of the subjective approach as it does not rely on an individual evaluation but on the distribution observed in the labor market. However, the use of a range of one standard deviation was criticized as arbitrary choice (Bauer, 2002) and the method was shown to be prone to outliers (Kiker, Santos, and de Oliveira, 1997). Kiker, Santos, and de Oliveira (1997) also used the empirical method but instead of relying on the mean value, they used the modal value within an occupation. This approach keeps the advantages of the approach by Verdugo and Verdugo (1989) without being prone to its abovementioned weaknesses. Within the course of this analysis, this modal value approach is applied.

The analysis starts with classical OLS wage regressions. However, they fail to provide causal effects because of unobserved heterogeneity. Applying OLS regressions requires the assumption that workers only differ by their observed characteristics, which is highly doubtful. As argued in the more recent literature on mismatch (Bauer, 2002; Korpi and Tahlin, 2009), unobserved influences such as intelligence, productivity and motivation are important factors when analyzing mismatch.

This analysis uses three different approaches to tackle the problem of unobserved heterogeneity. Bauer (2002) points out that using panel data, it is possible to estimate fixed effects regressions to control for unobserved influences. This approach is reproduced in this study. The wage regressions presented above remain the same, however, using the panel nature of the SOEP, individual fixed effects are controlled for. While this strategy controls for unobserved heterogeneity, it introduces another problem. Only individuals changing their educational match within the observation period can be used in fixed effects regressions, which implies that many mismatch observations cannot be regarded within the analysis. To account for this problem, an IV approach is used. Within the mismatch context, this was done by Korpi and Tahlin (2009), who also point to the difficulty of using an IV approach in the OMU framework: All *three* education variables, overeducation, undereducation and the required education must be instrumented. Hence, at least three instruments have to be found that fulfill the criteria of instrumental variables; all instruments have to be correlated with the instrumented variable (relevance) while they must not affect the outcome variable over a different channel (validity). Korpi and Tahlin (2009) apply four instruments in their analysis, all of them related to the respondents' youth: the number of siblings, place of residence, economic problems and family disruption. While the authors argue that these instruments are valid, they fail to fulfill the relevance criterion.

Applying instrumental variable approaches to account for unobserved heterogeneity in education studies is not a new idea. Angrist and Krueger (1991) use information on the quarter of birth on US census data. Card (1993) famously uses the proximity of the place of residence to the nearest college, which was later shown to be a weak instrument (for a discussion, see Harmon, Oosterbeek, and Walker, 2003). As already pointed out, unlike studies on the effects of education in general, this analysis requires three different instruments to cover all three possible matches.

The first, also used by Korpi and Tahlin (2009), is in line with a study by Butcher and Case (1994), using the presence of siblings. I basically follow their reasoning by including the number of siblings as an instrument for the educational match. The number of siblings is negatively correlated with the years of education. This is due to a split of parental support and expectations on several children. Parents can be assumed to lay a strong focus on the educational career of single children and this focus shifts, the higher the number of siblings, causing a lack of parental support during education (Butcher and Case, 1994). The data support this, as the number of siblings is negatively correlated with overeducation and the achieved education, while it is positively correlated with the case of undereducation.

The second instrument applied is also in line with studies stressing the family background, like Harmon and Walker (2000). Following Korpi and Tahlin (2009), an indicator for family disruption is applied. The number of years living without the biological parents until the age of 16 is used as an instrument, following a similar reasoning as for the number of siblings. The more time spent living with the biological parents, the higher the support and achievements in the educational career. For the third instrument, I use macro changes for exogenous variation. Often used are schooling reforms (e.g. Harmon and Walker, 1995, 1999; Pons and Gonzalo, 2002). However, as the data applied here cover a long time and observations from all age groups, schooling reforms cannot be used. Instead, I apply labor market conditions at the respondents' age of 15. This is a time in which individuals decide (1) to stay in the academic track; (2) to leave for the labor market aiming at a more applied vocational education; (3) not to obtain any further education[1].

A high unemployment rate at this time is likely to influence individuals to stay in the schooling system as an outside option to entering the labor market, which does not offer good opportunities at that time. The data do indeed support this hypothesis. While these three variables all have an influence on the educational decision and thereby on the educational match, they do not have other wage effects. All three instruments used in the analysis refer to living conditions while respondents are in education. This is the theoretical basis for the effect the instruments have on the education variables, which is also verified by statistical tests. While the amount of education is clearly influenced by these living conditions, there is no other channel to be found over which the instruments affect the hourly wages in later life.

Fixed effects regressions have the abovementioned weakness of missing individuals not changing their educational match. This problem is avoided by IV regressions. As a robustness check, a third method of dealing with the problem of unobserved heterogeneity is applied. In this approach, innate ability, which is unobserved to the researcher, is directly modeled in the regressions by including a proxy. In the OMU framework, Korpi and Tahlin (2009) make an approach to do this by including measures of health and verbal ability to capture this dimension. However, including proxies does not change their results significantly. In the SOEP, there is no direct measure of ability included. I therefore use data from the International Adult Literacy Survey (IALS) by the OECD. This dataset includes information on work-related reading, writing and math skills. These measure ability of individuals from a more applied perspective than the formal educational attainment and can be used as proxies for innate ability.

Applying the FE approach, I can reproduce the findings by Bauer (2002). In a second step, the use of instruments for the educational match avoids the problem of fixed effects regressions in education studies and refutes previous results. To test the stability of these new findings for Germany, I use data from the International Adult Literacy Survey (IALS) and include reading, writing and math skills as proxy for innate ability.

[1] Data on the yearly unemployment rate is provided by the Federal Statistical Office of Germany. For further information, see `https://www.destatis.de/EN/AboutUs/AboutUs.html`

4 Data

The data used in the first part of this analysis is from the German Socio-Economic Panel (SOEP), one of the longest running representative panel-datasets in Europe. Established in 1984, it covers more than 20,000 individuals per year and is representative of the German population (Wagner, Frick, and Schupp, 2007)[2].

The sample is restricted to working individuals between 18 and 65 and covers the years 1991 - 2011. I exclude individuals who lived in Eastern Germany before German reunification as well as immigrants who came to Germany after their 10th birthday. These restrictions ensure that respondents were educated in the same educational system and results are not driven by a different perception of educational titles from other countries. This proceeding replicates Bauer (2002), who also uses SOEP data for his analysis of educational mismatch.

Table 1: Descriptive Statistics – Control Variables

	Men		Women	
	Mean	Std.Dev.	Mean	Std.Dev.
Hourly Income	9.81	(6.31)	7.29	(4.85)
Age	40.35	(11.56)	39.49	(11.44)
Cohort	4.76	(1.23)	4.90	(1.20)
Nights in Hospital	0.75	(4.94)	0.76	(4.56)
Children	0.73	(0.95)	0.61	(0.87)
Fulltime	0.94	(0.23)	0.52	(0.50)
Tenure	12.17	(10.81)	9.16	(9.14)
Married	0.67	(0.47)	0.62	(0.48)
N	55415		45529	

Note: Authors' calculations based on the SOEP.

Table 1 presents descriptive statistics for the variables used in the wage regressions. The first column presents statistics for men, the second for women. Average hourly wages for males are €9.81, for females €7.29. The mean age is around 40 for men, 39.5 for females. Men in the sample have 0.72 children, women 0.6 on average. This can be explained by a lower labor market participation rate for mothers in Germany. The health control shows nearly equal values for men and women. Labor market specific controls show that tenure is about three years higher for men, at about 12 years. Men are most likely fulltime employed (94%), while a little more than half of the women in the sample work fulltime.

In Table 2, descriptive statistics for the main interest variables in the OMU framework can be found. To derive results which show the situation in Germany as precise

[2] All data were extracted using the Stata add-on PanelWhiz, written by Prof. Dr. John P. Haisken-DeNew (Haisken-DeNew and Hahn, 2010).

Table 2: Descriptive Statistics – Educational Mismatch

	Men		Women	
	Mean	Std.Dev.	Mean	Std.Dev.
OE Mode	2.16	(2.83)	1.59	(2.40)
Educ Mode	14.59	(2.27)	14.48	(1.98)
UE Mode	0.43	(1.04)	0.58	(1.06)
N	55415		45529	

Note: Authors' calculations based on the SOEP.

as possible, I avoid using a standard education variable in the SOEP, which offers mapped information on the years of education.[3] This variable assigns each individual the years of education typically necessary to obtain the highest achieved degree. For example, leaving education after the A levels (*Abitur*) means having 13 years of education, a vocational education means further 1.5 - 2 years, a university education 5 further years. While this is a useful tool for many applications, it does not fully meet the requirements of this study. Especially for higher educated individuals, this leads to measurement problems. For the mapped variable, the standard period of studies is used. This, however, is likely to be extended, the higher the educational degree. For a vocational education, which follows a 2-pillar strategy in Germany with school participation and an applied education in the labor market, there are different possibilities regarding the length of education. Degrees from universities and polytechnics are in many cases not achieved within the standard period. This can lead to an underestimation of the attained years of education and thereby to an incorrect measure of over- and undereducation. Therefore, I use the spell data in the SOEP. Education spells within the sample period are directly observed. For education spells before entering the sample, I use data from the biography questionnaire. Respondents entering the SOEP fill in a questionnaire stating their main activity (e.g. *in education, employed, unemployed*) for each year since their 15th birthday. This information is used to derive the overall years of education respondents have really spent in education[4].

Table 3: Wage Regressions - Full Sample

	Log. Hourly Income		
	(OLS)	(FE)	(IV)
OE Mode	0.019***	0.047***	-0.049
	(0.001)	(0.004)	(0.047)
Educ Mode	0.075***	0.053***	0.107***
	(0.001)	(0.004)	(0.020)
UE Mode	-0.046***	-0.058***	-0.383***
	(0.001)	(0.005)	(0.118)
Female	-0.222***	–	-0.213***
	(0.003)		(0.018)
Constant	Yes	Yes	Yes
Standard Controls	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
N	100944	100944	100944

Note: Robust standard errors in parentheses. *,** and *** denote significance level of 10%, 5% and 1% respectively. Estimations based on SOEP data 1991 - 2011. Excluded instruments are the national unemployment ate at the age of 15; Number of siblings; Years spent without natural parents during childhood.

5 Results

Table 3 shows the results of the wage regressions for the pooled sample. The first column shows the results from OLS wage regressions using the OMU framework. The classical OMU finding, i.e. positive returns to overeducation, which are lower than the positive returns to required education, and negative returns to undereducation, is remarkably stable across countries and datasets (Hartog, 2000). It is therefore not surprising that this result can also be found here. The OLS regressions show positive returns to overeducation of 2% per year of overeducation. The returns to the years of required education are about 7.5% and significantly higher. The negative effect of undereducation is nearly 5% per year. All coefficients are highly significant and in line with the previous literature.

Applying fixed effects changes the results drastically. The coefficient for the required education within an occupation becomes smaller (0.053). The coefficients for years of over- and undereducation are larger than in the OLS regressions. The wage benefit for a year of overeducation is 4.7% and thereby nearly the same as the one for years of required education. Hence, the difference between these two influences nearly vanishes here. The coefficient for years of undereducation is also larger than in the OLS case, showing a wage penalty of more than 7% per year. All coefficients remain highly significant. These findings are similar to the fixed effects results by Bauer (2002), who also uses the SOEP, but for a different observation period.

These results hint at the validity of the assumption that there is compensation taking

[3] A detailed documentation of the data properties is offered by the data provider, the DIW (see http://www.diw.de). A documentation of the mapping of years of education is provided in the SOEP documentation by Anger (2011).

[4] Robustness checks using the mapped years of education variable shows that most results are qualitatively but not always quantitatively comparable. For tables, see Appendix.

place, at least for overeducated workers. When controlling for unobserved heterogeneity using individual fixed effects, the coefficients for years of required education and overeducation become closer to each other, as expected in this theory. However, the FE method to control for unobserved factors has some weaknesses in this context. Only individuals changing their match are observed as all non-changers are time-invariant and therefore not included in the panel model. Changing the match means (1) changing the job; (2) an overall shift of the requirements in the job; (3) going back to education. All of these cases are rather special, while many cases in which individuals do not change their match over the sample period cannot be observed. Also, the concentration on those changers is problematic as there might be other unobserved factors underlying these changes.

The third column of Table 3 shows the results for the IV regressions. Here, the difference to the OLS results is even larger. Returns to years of overeducation are negative but insignificant. This implies that the pattern clearly differs from the OLS and FE findings. The returns to years of required education are higher than in the previous regressions (0.107). The coefficient for years of undereducation is much larger than in the previous regressions and still highly significant. The effect is noticeably large, which can be explained by two arguments: (1) IV regressions are less precise than OLS regressions (Wooldridge, 2000), which leads to larger confidence intervals and makes the point estimates less meaningful; (2) while the first-stage regressions[5] and Angrist-Pischke multivariate F test of excluded instruments (Angrist and Pischke, 2008) suggest that the instruments are strong for all three education components, the value for the undereducation case is much smaller (an F-value of 20) than for the other two match variables. There is a significant correlation of the instruments with the endogenous variable to be seen but this is not as strong as in the overeducation and required education cases. For these reasons, I do not interpret the magnitude of the undereducation point estimates.

The results of the FE as well as the IV regressions point at a bias in the OLS results. However, the direction of the bias is ambiguous. While the FE results suggest that OLS results are biased downwards, IV results suggest that the overeducation coefficient is overestimated, leading to very different conclusions. While the FE findings back the theory of human capital compensation, the IV results hint at a structural problem in the German labor market. Only years of education really required in a job pay off, while any additional education does not. This means that education over and above what is required is not productive in the labor market, otherwise employers would pay for it. At the same time, overeducation is a widespread phenomenon.

[5] The results of the first-stage regressions can be found in the Appendix.

Hence, education is attained which then remains unproductive in the labor market. This means a huge waste of public resources as well as individual effort in the educational sector. As the two explanations differ gravely, I shed further light by splitting the sample to see whether there are different patterns for men and women.

Table 4: Wage Regressions - Male Sample

	Log. Hourly Income		
	(OLS)	(FE)	(IV)
OE Mode	0.019***	0.041***	-0.002
	(0.001)	(0.005)	(0.035)
Educ Mode	0.066***	0.046***	0.119***
	(0.001)	(0.005)	(0.017)
UE Mode	-0.043***	-0.045***	-0.164
	(0.002)	(0.006)	(0.119)
Constant	Yes	Yes	Yes
Standard Controls	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
N	55415	55415	55415

Note: Robust standard errors in parentheses. *,** and *** denote significance level of 10%, 5% and 1% respectively. Estimations based on SOEP data 1991 - 2011. Excluded instruments are the national unemployment rate at the age of 15; Number of siblings; Years spent without natural parents during childhood.

Table 4 shows the results for the male sample and the pattern found for the pooled sample is reproduced. The OLS results in column 1 reveal the classical pattern of OMU studies. In the FE regression, the coefficients for required education and overeducation become close to each other while the undereducation coefficient remains mainly unchanged. IV results (column 3) show that the overeducation coefficient becomes negative and insignificant, while the required education effect becomes larger and remains statistically significant. Undereducation remains negative and also becomes larger, however, it is statistically insignificant. Hence, the findings for the male sample are nearly the same as for the pooled one.

Table 5: Wage Regressions - Female Sample

	Log. Hourly Income		
	(OLS)	(FE)	(IV)
OE Mode	0.015***	0.049***	-0.258
	(0.001)	(0.006)	(0.185)
Educ Mode	0.083***	0.056***	0.135**
	(0.001)	(0.007)	(0.057)
UE Mode	-0.047***	-0.072***	-0.686**
	(0.002)	(0.008)	(0.280)
Constant	Yes	Yes	Yes
Standard Controls	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
N	45529	45529	45529

Note: Robust standard errors in parentheses. *,** and *** denote significance level of 10%, 5% and 1% respectively. Estimations based on SOEP data 1991 - 2011. Excluded instruments are the national unemployment rate at the age of 15; Number of siblings; Years spent without natural parents during childhood.

Table 5 shows results for female respondents. Once again, the pattern found before is confirmed. In the OLS regressions, required education has a larger effect than overeducation while this changes in the FE regression. The overeducation effect becomes larger, the required education effect becomes smaller and both become closer to each other. While this finding again seems to confirm the compensation hypothesis, the IV results show a very different pattern, rather supporting the theory of a general matching problem between the educational supply and demand.

All in all, the results of the wage regressions show some remarkable findings. The OLS wage regressions reveal the expected results. The FE results do not confirm this general finding completely as the difference between overeducation and required education is not found anymore. This is very similar to the findings by Bauer (2002) and supports the theory explaining educational mismatch as a form of compensation for other forms of ability. The IV results, however, reject this hypothesis by showing that the effect of matched years of education becomes larger than in the other regressions while the overeducation effect becomes insignificant, the point estimate even negative. This backs the theory of an allocation problem as many individuals acquire education which then is not used in the labor market.

As the FE and IV regressions show different results, I expand the empirical strategy and apply a third strategy, to ensure the robustness of the IV findings. Here, I proxy ability with reading, writing and math skills and include these in the OMU regressions. Unfortunately, I cannot achieve this by using data from the SOEP because there is no ability measure included. Instead, I use data from the International Adult Literacy Survey. This is a joint project of the OECD and Statistics Canada[6]. In 1994, representative samples from European and Northern American countries were interviewed with the aim of getting a comprehensive picture of skills among adults, exceeding the measure of formal educational attainment. These include numeracy as well as literacy proficiency. Using the observations from this dataset from Germany, it is possible to include skill measures as proxies for ability in the OMU framework directly. Unlike the SOEP, the IALS data is a cross-section and not a panel. However, including the skill measures directly, the panel dimension is expendable for this step of the analysis. The dataset does not include a continuous wage variable but wage quintiles. This changes the econometric approach. I apply three different models to guarantee that the results are not driven by the choice of the model. Firstly, I estimate a linear OLS model with the wage as 5-digit variable. As the dependent variable is the wage and not a classical categorical variable, the assumption of linearity is not supposed to cause problems. However, to ensure this, I estimate an ordered logit as second model. In a third step, I apply an interval regression.[7] As the econometric framework is different, the data set is much smaller and I do not have the whole set of standard wage-regression control variables, I do not interpret coefficients quantitatively. However, they can show the direction of the

[6] A detailed description of the data can be found in Murray, Kirsch, and Jenkins (1998).
[7] The income intervals are generated using the SOEP income quintiles.

findings.

Table 6: Descriptive Statistics – IALS Data

	IALS data	
	Mean	Std.Dev.
Female	0.47	(0.50)
OE Mode	1.42	(2.28)
UE Mode	0.75	(1.61)
Educ Mode	10.93	(2.88)
Employer unchanged	0.80	(0.40)
Fulltime	0.68	(0.47)
Math Skills	1.71	(0.64)
Reading Skills	1.49	(0.56)
Writing Skills	1.58	(0.60)
N	1025	

Note: Authors' calculations based on the IALS.

Table 6 shows the descriptive statistics of the IALS data.[8] The sample is evenly distributed by sex. Required education is around 11 years, which is lower than in the SOEP sample. This is due to the construction of the variable, as I cannot use a measure as precise as the one derived from the SOEP spell data. The years of required education are closer to the mapped SOEP variable of years of education. 80% of the respondents have not changed their employer in the last 12 month, which serves as a proxy for tenure. The numeracy and literacy skills are self-assessed on a scale from 1 (*excellent*) to 4 (*poor*). Mean literacy skills are a little better than numeracy skills, with reading skills closest to *excellent*.

Table 7 shows the results of the IALS regressions without including ability controls. The general pattern of the OMU literature and the previous OLS regressions is reproduced. Overeducation has a positive significant coefficient, required education is also positively significant and larger. Undereducation does not have a negative coefficient here but it still has the lowest point estimate and is insignificant.

Table 8 presents the results controlling for numeracy and literacy skills. In all three models, the effect of overeducation becomes smaller and loses significance, in the OLS and ordered logit models it stays weakly significant at the 10% level, in the interval regression it becomes statistically insignificant. The coefficients of required education remain positive and highly significant, so the gap between the coefficients of overeducation and required education widens. This differs from the previous FE findings but is in line with the IV findings.

[8] I have not split the IALS data further into a male and a female sample. This is due to the significantly lower number of observations as compared to the SOEP, with a total of about 1000 observations of working individuals in the IALS. As the SOEP regressions show, the sample split does not change the findings gravely. Robustness checks with a split sample of the IALS data show that the same is true here.

Table 7: Wage Regressions without Abitily Controls - IALS Data

	Wage		
	(OLS)	(OLogit)	(Interval)
OE Mode	0.040**	0.073***	0.013**
	(0.016)	(0.028)	(0.006)
Educ Mode	0.100***	0.185***	0.040***
	(0.015)	(0.027)	(0.006)
UE Mode	0.024	0.029	0.008
	(0.028)	(0.049)	(0.011)
Constant	1.476***	–	6.457***
	(0.175)		(0.070)
cut1	–	0.476	–
		(0.305)	
cut2	–	2.182***	–
		(0.306)	
cut3	–	3.403***	–
		(0.314)	
cut4	–	4.741***	–
		(0.332)	
lnsigma	–	–	-0.927***
			(0.030)
Standard Controls	Yes	Yes	Yes
N	1025	1025	1025

Note: Robust standard errors in parentheses. *,** and *** denote significance level of 10%, 5% and 1% respectively. Estimations based on IALS data.

The results from the IALS sample show that the classical OMU findings can be reproduced without ability controls. When controlling for ability directly, the results change significantly into the direction of the IV results for the SOEP sample. This means they reject the human capital compensation theory. Instead, they show an allocation problem in the labor market with many individuals spending a long time in education, accumulating human capital which then is not used productively in the labor market.

6 Conclusion

This paper analyses the causal effects of educational mismatch on wages in Germany using data from the SOEP and the IALS. Educational mismatch is defined as a situation in which individuals have more or less formal education than the modal value within an occupation (empirical definition of mismatch). OLS results confirm the previous findings of this field of literature. As more recent studies on the wage effects of over-/undereducation have found, these results might be biased due to unobserved heterogeneity (Bauer, 2002; Korpi and Tahlin, 2009). Taking this into account, I apply a fixed effects approach, an IV approach and use data from the IALS to model skills as proxies for ability in the regression framework directly. Results from the FE regressions confirm the results by Bauer (2002) and suggest the validity of the human capital compensation theory. According to this theory, overeducated workers compensate for lower innate ability. With fixed effects regressions, I cannot observe non-changers in the context of education, thus I extend the econometric strategy.

Table 8: Wage Regressions - IALS Data

	Wage		
	(OLS)	(Ologit)	(Interval)
OE Mode	0.028*	0.050*	0.009
	(0.016)	(0.028)	(0.006)
Educ Mode	0.087***	0.162***	0.035***
	(0.016)	(0.028)	(0.006)
UE Mode	0.032	0.044	0.011
	(0.028)	(0.049)	(0.011)
Math Skills	-0.142**	-0.275**	-0.057**
	(0.068)	(0.114)	(0.026)
Writing Skills	-0.148	-0.261	-0.061
	(0.098)	(0.162)	(0.037)
Reading Skills	0.061	0.091	0.034
	(0.100)	(0.166)	(0.038)
Constant	2.046***	–	6.666***
	(0.242)		(0.095)
cut1	–	-0.616	–
		(0.412)	
cut2	–	1.109***	–
		(0.409)	
cut3	–	2.348***	–
		(0.412)	
cut4	–	3.706***	–
		(0.422)	
lnsigma	–	–	-0.934***
			(0.030)
Standard Controls	Yes	Yes	Yes
N	1025	1025	1025

Note: Robust standard errors in parentheses. *,** and *** denote significance level of 10%, 5% and 1% respectively. Estimations based on IALS data.

Using an IV approach, I account for endogeneity of the education variables without the restriction of time-invariance of the educational match. These results reject the compensation hypothesis, whereby over- and undereducation become insignificant and only required education has a positive earnings effect. This finding is supported by regressions using data from the IALS, which includes measures of numeracy and literacy skills among adults and can be used to proxy ability in the OMU framework. The regressions show the robustness of the IV regressions, with required education as only significantly positive influence on earnings.

This study contributes to the general discussion about the causal effects of educational mismatch and the mechanisms behind it. The results obtained here refute the compensation theory, as it is rejected by all models but the FE panel model. However, there is a different pattern visible in the data. Eliminating unobserved heterogeneity, only the effect of required education remains positive and significant.

This study shows that there is a problem in the German educational system and its link to the labor market. Results show that there is hardly a positive causal effect of overeducation to be found, which means that this additional human capital is unproductive. Overeducation is a common feature of the German labor market with more than 50% of employees in the situation of an educational mismatch. If the additional education is mainly unproductive, this is a massive waste of public resources. Individuals could enter the labor market earlier instead of spending further years in education. While being in education, individuals face foregone wages. This means the direct loss of wages for several years[9] and future pension payments are reduced as the time of contributing to the pension fund is lowered.[10] If people are overeducated later, investments in human capital do not fully pay off. In additions to the individual costs of overeducation, there are also public costs to be regarded. Education is mainly financed by public expenditure in Germany.[11] Individuals do not have to pay for their education directly – besides the abovementioned foregone wages – and as a result, they structurally overinvest in education.

The introduction of tuition fees could lead to a better allocation of resources. When not only foregone wages had to be considered but there was a direct price to be paid for education, individuals would consider their investment in education more carefully. Having a clear idea of the occupation they want to work in, individuals could choose their amount of education more purposefully and the likelihood of overeducation would decrease. Due to this, the educational system could be slimmed down, relieving public coffers. At the same time, public expenditures could be further reduced as those individuals staying in education would pay for it. As it becomes more likely to find an optimal job match in the labor market, higher wages compensate for these expenditures.

Overeducation plays an important role in Germany showing that there is no general skill shortage in the labor market but an oversupply. Of course this does not rule out the possibility of a shortage of high-skilled individuals in certain fields, but this is not true for the labor market in general. Overall, the allocation does not work perfectly. Over the past decades, skill demand has risen significantly with the technological development. At the moment, it seems as if the labor market is saturated and the educational system systematically produces overskilled individuals.

[9] Mean overeducation in this sample is about 2.2 years for men and 1.6 years for women.

[10] The amount of money paid after retirement depends on the time in the labor market.

[11] In the early 2000s, some federal states introduced tuition fees for higher education of €1,000 per year. Altough this amount is very moderate compared to tuition fees in most other industrialized countries, the tuition fees were abolished after a few years due to severe student protests.

References

Alba-Ramirez, A. (1993): "Mismatch in the Spanish Labor Market: Overeducation?," *Journal of Human Resources*, pp. 259–278.

Anger, S. (2011): "Documentation PGEN - Person-related Status and Generated Variables," SOEP documentation.

Angrist, J. D., and A. B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *The Quarterly Journal of Economics*, 106(4), 979–1014.

Angrist, J. D., and J.-S. Pischke (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 1 edn.

Battu, H., C. R. Belfield, and P. J. Sloane (1999): "Overeducation Among Graduates: A Cohort View," *Education Economics*, 7(1), 21–38.

Bauer, T. K. (2002): "Educational Mismatch and Wages: A Panel Analysis," *Economics of Education Review*, 21(3), 221 – 229.

Butcher, K. F., and A. Case (1994): "The Effect of Sibling Sex Composition on Women's Education and Earnings," *The Quarterly Journal of Economics*, 109(3), 531–563.

Card, D. (1993): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," Discussion paper, National Bureau of Economic Research.

Duncan, G. J., and S. D. Hoffman (1981): "The Incidence and Wage Effects of Overeducation," *Economics of Education Review*, 1(1), 75 – 86.

Groeneveld, S., and J. Hartog (2004): "Overeducation, Wages and Promotions within the Firm," *Labour Economics*, 11(6), 701 – 714.

Haisken-DeNew, J. P., and M. H. Hahn (2010): "PanelWhiz: Efficient Data Extraction of Complex Panel Data Sets - An Example Using the German SOEP," *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 130(4), 643–654.

Harmon, C., H. Oosterbeek, and I. Walker (2003): "The Returns to Education: Microeconomics," *Journal of Economic Surveys*, 17(2), 115–156.

Harmon, C., and I. Walker (1995): "Estimates of the Economic Return to Schooling for the United Kingdom," *The American Economic Review*, 85(5), 1278–1286.

——— (1999): "The Marginal and Average Returns to Schooling in the UK," *European Economic Review*, 43(4-6), 879 – 887.

——— (2000): "The Returns to the Quantity and Quality of Education: Evidence for Men in England and Wales," *Economica*, 67(265), 19–35.

Hartog, J. (2000): "Over-education and Earnings: Where Are We, Where Should We Go?," *Economics of Education Review*, 19(2), 131 – 147.

Hartog, J., and H. Oosterbeek (1988): "Education, Allocation and Earnings in the Netherlands: Overschooling?," *Economics of Education Review*, 7(2), 185–194.

Kiker, B., M. C. Santos, and M. de Oliveira (1997): "Overeducation and undereducation: Evidence for Portugal," *Economics of Education Review*, 16(2), 111 – 125.

Korpi, T., and M. Tahlin (2009): "Educational Mismatch, Wages, and Wage Growth: Overeducation in Sweden, 1974 - 2000," *Labour Economics*, 16(2), 183 – 193.

McGoldrick, K., and J. Robst (1996): "Gender Differences in Overeducation: A Test of the Theory of Differential Overqualification," *The American Economic Review*, 86(2), pp. 280–284.

Mincer, J. (1970): "The Distribution of Labor Incomes: A Survey With Special Reference to the Human Capital Approach," *Journal of Economic Literature*, 8, 1–26.

Murray, T. S., I. S. Kirsch, and L. B. Jenkins (1998): *Adult literacy in OECD countries: Technical report on the first International Adult Literacy Survey*. US Department of Education, Office of Educational Research and Improvement.

Pons, E., and M. T. Gonzalo (2002): "Returns to Schooling in Spain: How Reliable are Instrumental Variable Estimates?," *Labour*, 16(4), 747–770.

Rubb, S. (2003): "Overeducation in the Labor Market: A Comment and Re-analysis of a Meta-analysis," *Economics of Education Review*, 22(6), 621 – 629.

Rumberger, R. W. (1987): "The Impact of Surplus Schooling on Productivity and Earnings.," *Journal of Human Resources*, 22(1), 24 – 50.

Sicherman, N. (1991): "Overeducation in the Labor Market," *Journal of Labor Economics*, 9(2), pp. 101–122.

Sloane, P. J., H. Battu, and P. T. Seaman (1999): "Overeducation, Undereducation and the British Labour Market," *Applied Economics*, 31(11), 1437–1453.

Verdugo, R. R., and N. T. Verdugo (1989): "The Impact of Surplus Schooling on Earnings," *The Journal of Human Resources*, 24(4), 629–643.

Wagner, G. G., J. R. Frick, and J. Schupp (2007): "The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements," *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 127(1), 139–169.

Wooldridge, J. (2000): "Introductory Econometrics, A Modern Approach, 2003," *South-Western, Thomson Learning*.

Appendix

Table 9: First Stage Regressions – Pooled Sample

	Education Components		
	(OE)	(Educ.Mode)	(UE)
No. of Siblings	-0.128***	-0.129***	0.037***
	(0.004)	(0.004)	(0.002)
Yrs without Bio. Par.	0.021***	0.006	0.002
	(0.005)	(0.004)	(0.002)
UE Rate at Age 15	0.063***	0.004	-0.025***
	(0.004)	(0.003)	(0.002)
Age	0.333***	0.089***	-0.074***
	(0.006)	(0.005)	(0.003)
Age Squarred	-0.004***	-0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Cohort	-0.126***	0.046**	0.021*
	(0.026)	(0.022)	(0.011)
Nights in Hospital	-0.004**	-0.006***	0.001*
	(0.002)	(0.001)	(0.001)
Children	-0.147***	0.099***	0.021***
	(0.010)	(0.009)	(0.004)
Fulltime	0.107***	0.484***	-0.029***
	(0.024)	(0.019)	(0.009)
Tenure	-0.052***	0.002**	0.009***
	(0.001)	(0.001)	(0.000)
Married	-0.524***	-0.155***	0.070***
	(0.024)	(0.018)	(0.008)
Female	-0.740***	0.109***	0.187***
	(0.020)	(0.016)	(0.007)
Constant	-3.577***	11.432***	1.645***
	(0.229)	(0.192)	(0.100)
Year dummies	Yes	Yes	Yes
N	100944	100944	100944

Note: Robust standard errors in parentheses. *,** and *** denote significance level of 10%, 5% and 1% respectively. Estimations based on SOEP data 2000 - 2011. First stage regressions of IV wage regressions presented in this analysis.

Table 10: First Stage Regressions – Male Sample

	Education Components		
	(OE)	(Educ.Mode)	(UE)
No. of Siblings	-0.154***	-0.135***	0.033***
	(0.006)	(0.005)	(0.003)
Yrs without Bio. Par.	0.020***	-0.000	0.005**
	(0.007)	(0.005)	(0.003)
UE Rate at Age 15	0.067***	-0.004	-0.020***
	(0.006)	(0.004)	(0.002)
Age	0.370***	0.102***	-0.067***
	(0.009)	(0.007)	(0.004)
Age Squarred	-0.004***	-0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Cohort	-0.060	0.033	0.027*
	(0.038)	(0.031)	(0.015)
Nights in Hospital	-0.005*	-0.012***	0.001
	(0.003)	(0.002)	(0.001)
Children	-0.107***	0.110***	0.021***
	(0.014)	(0.012)	(0.006)
Fulltime	-1.055***	-0.208***	-0.004
	(0.069)	(0.045)	(0.018)
Tenure	-0.060***	-0.018***	0.009***
	(0.002)	(0.001)	(0.001)
Married	-0.406***	-0.095***	0.065***
	(0.037)	(0.027)	(0.011)
o.Female	0.000	0.000	0.000
	(.)	(.)	(.)
Constant	-3.850***	11.776***	1.416***
	(0.330)	(0.271)	(0.133)
Year dummies	Yes	Yes	Yes
N	55415	55415	55415

Note: Robust standard errors in parentheses. *,** and *** denote significance level of 10%, 5% and 1% respectively. Estimations based on SOEP data 2000 - 2011. First stage regressions of IV wage regressions presented in this analysis.

Table 11: First Stage Regressions – Female Sample

	Education Components		
	(OE)	(Educ.Mode)	(UE)
No. of Siblings	-0.086***	-0.113***	0.041***
	(0.006)	(0.005)	(0.003)
Yrs without Bio. Par.	0.017**	0.009*	-0.001
	(0.007)	(0.006)	(0.003)
UE Rate at Age 15	0.043***	0.001	-0.029***
	(0.005)	(0.004)	(0.003)
Age	0.326***	0.098***	-0.085***
	(0.008)	(0.007)	(0.004)
Age Squarred	-0.004***	-0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Cohort	-0.198***	0.061**	0.011
	(0.035)	(0.030)	(0.016)
Nights in Hospital	-0.004*	0.001	0.002*
	(0.002)	(0.002)	(0.001)
Children	-0.187***	0.096***	0.027***
	(0.015)	(0.013)	(0.007)
Fulltime	0.263***	0.526***	-0.021*
	(0.027)	(0.022)	(0.011)
Tenure	-0.044***	0.027***	0.009***
	(0.001)	(0.001)	(0.001)
Married	-0.618***	-0.213***	0.080***
	(0.031)	(0.024)	(0.012)
o.Female	0.000	0.000	0.000
	(.)	(.)	(.)
Constant	-3.463***	11.515***	2.123***
	(0.306)	(0.265)	(0.152)
Year dummies	Yes	Yes	Yes
N	45529	45529	45529

Note: Robust standard errors in parentheses. *,** and *** denote significance level of 10%, 5% and 1% respectively. Estimations based on SOEP data 2000 - 2011. First stage regressions of IV wage regressions presented in this analysis.

Table 12: Wage Regressions - Full Sample - Mapped Years of Educ.

	Log. Hourly Income		
	(OLS)	(FE)	(IV)
OE Mode	0.049***	0.055***	-0.061
	(0.001)	(0.004)	(0.451)
Educ Mode	0.079***	0.062***	0.057
	(0.001)	(0.004)	(0.100)
UE Mode	-0.055***	-0.061***	-0.655***
	(0.001)	(0.004)	(0.245)
Female	-0.239***	–	-0.330***
	(0.003)		(0.062)
Constant	Yes	Yes	Yes
Standard Controls	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
N	100074	100074	100074

Note: Robust standard errors in parentheses. *,** and *** denote significance level of 10%, 5% and 1% respectively. Estimations based on SOEP data 1991 - 2011. Excluded instruments are the national unemployment rate at the age of 15; Number of siblings; Years spent without natural parents during childhood.

Table 13: Wage Regressions - Male Sample - Mapped Years of Educ.

	Log. Hourly Income		
	(OLS)	(FE)	(IV)
OE Mode	0.054***	0.052***	-0.005
	(0.002)	(0.005)	(0.083)
Educ Mode	0.072***	0.058***	0.093***
	(0.001)	(0.004)	(0.022)
UE Mode	-0.048***	-0.057***	-0.217***
	(0.001)	(0.004)	(0.064)
Constant	Yes	Yes	Yes
Standard Controls	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
N	54721	54721	54721

Note: Robust standard errors in parentheses. *,** and *** denote significance level of 10%, 5% and 1% respectively. Estimations based on SOEP data 1991 - 2011. Excluded instruments are the national unemployment rate at the age of 15; Number of siblings; Years spent without natural parents during childhood.

Table 14: Wage Regressions - Female Sample - Mapped Years of Educ.

	Log. Hourly Income		
	(OLS)	(FE)	(IV)
OE Mode	0.038***	0.061***	1.063
	(0.002)	(0.006)	(1.966)
Educ Mode	0.084***	0.068***	-0.577
	(0.001)	(0.006)	(1.061)
UE Mode	-0.067***	-0.069***	-2.260
	(0.002)	(0.006)	(3.659)
Constant	Yes	Yes	Yes
Standard Controls	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
N	45353	45353	45353

Note: Robust standard errors in parentheses. *,** and *** denote significance level of 10%, 5% and 1% respectively. Estimations based on SOEP data 1991 - 2011. Excluded instruments are the national unemployment rate at the age of 15; Number of siblings; Years spent without natural parents during childhood.