

Saniter, Nils

**Working Paper**

## Estimating heterogeneous returns to education in Germany via conditional heteroskedasticity

SOEPpapers on Multidisciplinary Panel Data Research, No. 458

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Saniter, Nils (2012) : Estimating heterogeneous returns to education in Germany via conditional heteroskedasticity, SOEPpapers on Multidisciplinary Panel Data Research, No. 458, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<http://hdl.handle.net/10419/60189>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

## SOEPpapers

on Multidisciplinary Panel Data Research

# Estimating Heterogeneous Returns to Education in Germany via Conditional Heteroskedasticity

Nils Saniter

## **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:**

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)  
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)

# Estimating Heterogeneous Returns to Education in Germany via Conditional Heteroskedasticity

Nils Saniter \*

June 2012

## Abstract

In this paper I investigate the causal returns to education for different educational groups in Germany by employing a new method by Klein and Vella (2010) that bases identification on the presence of conditional heteroskedasticity. Compared to IV methods, key advantages of this approach are unbiased estimates in the absence of instruments and parameter interpretation that is not bounded to local average treatment effects. Using data from the German Socio-Economic Panel Study (SOEP) I find that the causal return to education is 8.5% for the entire sample, 2.3% for graduates from the basic school track and 11% for graduates from a higher school track. Across these groups the endogeneity bias in simple OLS regressions varies significantly. This confirms recent evidence in the literature on Germany. Various robustness checks support the findings.

**Keywords:** Return to education, wage equation, control function approach, second moment exclusion restriction

**JEL:** C3, I21, J31

---

\*DIW Berlin/Freie Universität Berlin, nsaniter@diw.de. I would like to thank Klaus F. Zimmermann, Thomas Siedler, Jörg Breitung, Francis Vella, Helmut Lütkepohl, Andrew Oswald, and Daniel S. Hamermesh as well as seminar participants at DIW Berlin and IZA Bonn for invaluable comments and suggestions. Furthermore, I am grateful to Simon Diehl for research assistance, to the Freie Universität Berlin for providing high-performance computing infrastructure, and to Adam Lederer for proofreading. The author bears the sole responsibility for the content of this paper.

# 1 Introduction

Ever since Mincer (1974) laid out the methodological foundation to estimate wage equations, a tremendous amount of work has been dedicated to finding the causal return to education. The causal return to education is the extra amount of wage income a randomly selected individual receives from an additional year of education. Knowing the causal return is important for policy makers. It directly informs about the utility of educational programs in terms of monetary payoffs for its beneficiaries. Estimating the causal returns, however, is not a trivial task: While simple wage regressions correctly produce correlations between, say, years of schooling and wages, they do not report causal returns to education as the schooling variable is likely to be endogenous due to omitted ability variables. Given the typical belief that the omitted ability variables influence wages and schooling in the same direction, simple OLS parameters are upward biased (Griliches 1977).<sup>1</sup> One well-established route to circumvent the endogeneity problem is to use instrument variable (IV) methods. While theoretically appealing, IV may not always be easily implemented in practice as it relies on the availability of valid and significant instruments. Moreover, the interpretation of IV parameters is often bounded to the *local average treatment effect* (LATE) along the lines of Imbens and Angrist (1994): The estimated coefficients represent causal effects only for the subsample of compliers, i.e. individuals who are actually affected by variations in the instrument. Inference on the *average treatment effect* (ATE), i.e. the entire population, is generally not valid. Therefore, different instruments typically produce different results and it requires case-specific judgement to determine which subgroup of the population the estimates are representing.

The guiding question of this paper is how years of education affect wages. I estimate the causal returns to education with ATE interpretation without using IV methods. Instead, I make use of a novel identification strategy by Klein and Vella (2010), which is realized with a control function approach. Identification relies on the nonlinearity of the control term induced by heteroskedasticity. That is, instead of first moment exclusion restrictions (mean), I make use of second moment exclusion restrictions (variance). Arguing that the presence of heteroskedasticity

---

<sup>1</sup>This disregards potential attenuation bias from measurement error.

is not limited to a subsample of the population, this allows me to estimate the true ATE. Data are taken from the German Socio- Economic Panel Study (SOEP) and results are produced for Germany for individuals of different educational groups – less well educated, better educated, and all individuals.

I will compare my results to three IV studies for Germany that use different instruments: Pischke and Wachter (2008) use the extension of compulsory schooling years across states and years as an instrument and find no returns to schooling. Their IV parameters are significantly smaller than their OLS coefficients, hinting at upward biased parameters in simple OLS regressions. Becker and Siebern-Thomas (2007) identify the returns to education using the urbanization of the place of childhood as an instrument to proxy the availability of higher secondary schools. They find downward biased OLS coefficients. Finally, Ichino and Winter-Ebmer (2004) instrument schooling with own and father’s World War II involvement and also find downward biased OLS estimates. While these results seem to contradict each other, they can be reconciled taking into account their LATE interpretation. Each IV result resembles the effect for the respective instrument-specific subpopulation of compliers. Arguably, for Pischke and Wachter (2008), these compliers are individuals with no or basic school education. For Becker and Siebern-Thomas (2007) they are individuals who are better educated. Finally, for Ichino and Winter-Ebmer (2004) they are individuals with any level of schooling.

The subsamples used in my study — low educated, better educated, and all individuals — resemble these complier groups. In fact, using these different subgroups, the three studies’ results can approximately be replicated using the approach by Klein and Vella (2010). This allows two suppositions: First, my method correctly identifies the causal education parameter. Second, the puzzle of seemingly conflicting evidence for Germany can be solved by accounting for IV’s LATE interpretation. To preview results, I find that the wage premium of one additional year of education in Germany is 8.5% for the whole sample. It is 11.5% for students with one of the two highest possible school diplomas, while for those who achieve a lesser high school diploma the return is 2.3%.

The remainder of this paper is organized as follows: Section 2 briefly describes the educational system in Germany. Section 3 reviews the relevant empirical lit-

erature. Section 4 presents the econometric model and the implementation of the estimator. Section 5 describes the data. Section 6 presents the empirical results. Discussions and robustness checks are provided in section 7. Finally, section 8 summarizes.

## 2 Education in Germany

In Germany, children typically start elementary school at the age of 6. After four years they move on to a secondary school and must choose between three school tracks that differ with respect to their curriculum and academic standards. The lower track (*Hauptschule*) is the least demanding one. Students finish this track after a total of 9 years of schooling. Besides basic academic content, this track contains various elements of vocational training. After finishing, graduates usually engage in an apprenticeship leading to a blue collar occupation. The middle track (*Realschule*) is more demanding than the lower track and ends after 10 years of schooling. Graduates of this track usually engage in an apprenticeship leading to a white collar occupation. The higher track (*Gymnasium*) is the most academic track. It ends after 13 years of schooling with the Abitur degree or after 12 years with the slightly less academic *Fachabitur* degree. While the former qualifies for university studies, the latter allows individuals to study at a polytechnic.<sup>2</sup>

The assignment to one of the school tracks is a combination of elementary school performance, teacher recommendation and parental choice. The exact mechanisms differ across states and years. Yet, everywhere in Germany and throughout history, school track choice is subject to considerable self-selection on ability and parental background (e.g. Dustmann 2004). The first two years of secondary school are often considered to be an orientation phase allowing for mobility between the tracks. If a student of a higher school track accomplishes the minimum number

---

<sup>2</sup>The emergence of new school forms and reductions in the duration of *Gymnasium* for *Abitur* graduates from 13 to 12 years in almost all German states are not relevant for my study, since my sample consists of early birth cohorts that were not affected by this change. Different years of schooling in the former German Democratic Republic do not play a role either as I focus only on West Germany. Besides the classical three school tracks, in some states comprehensive schools exist that comprise all the mentioned school tracks and may award all degrees. Numerically, however, comprehensive schools are not significant and receive no extra consideration in this study.

of years required for a lower school track, he or she can leave the school with the lower school degree (e.g. a student can finish *Realschule* after 9 instead of 10 years with a degree from *Hauptschule*).

What makes the German educational system different from many other countries is the important role of its vocational training (Gang and Zimmermann 2000). For graduates of the lower and middle school tracks it seems plausible that success in the labor market depends more on the type and quality of an apprenticeship than on the school degree itself (Pischke and Wachter 2008). Similarly, for individuals holding an *Abitur* degree, labor market success considerably improves with a university degree. Hence, a sole measure of years of schooling does not have sufficient explanatory power as to satisfactorily predict wages. This is why in my application I enhance the measure of schooling with information about vocational trainings, apprenticeships, polytechnics and university studies arriving at a measure of *total years of education*, as described in more detail in section 5.

### 3 Relevant literature

One major methodological avenue to estimate the returns to education free of endogeneity bias is instrumental variable methods. The IV approach builds upon the postulation that there exists an instrument variable that is correlated with the endogenous regressor but not with the error term. Causal effects can then be estimated using the exogenous variation of this variable. One often employed instrument with an arguably strong stance of validity is a compulsory schooling law change that brings about variation in the minimal number of schooling years across space and time. The pioneers of this strand of literature are Angrist and Krueger (1991), who find returns to schooling in the US labor market of 6-10% for different birth cohorts that lie well above their OLS estimates of 5-7%. This result is corroborated by Oreopoulos (2007), who calculates causal returns in the magnitude of 13% as compared to lower OLS returns of almost 8%. In contrast, Acemoglu and Angrist (2000) estimate a return to schooling of 10% and fail to find evidence for biased OLS estimates. For the UK, Harmon and Walker (1995) as well as Oreopoulos (2006) find roughly 15% higher earnings from one additional year of compulsory schooling, a result refuted by Devereux and Hart (2010), who calculate



only 3% returns on average. Similar studies exist for a whole range of other countries. Interestingly enough, many IV estimates range 20-40 percent higher than the corresponding OLS results (Card 1999). This hints at downward biased OLS estimates, which is counterintuitive given the above outlined interpretation of omitted ability variables. The most recent IV study on returns to education for Germany that uses school reform as an instrument is by Pischke and Wachter (2008) (PW henceforth). In the 1950s and 60s the duration of the basic track was extended from 8 to 9 years. The exact timing of the policy intervention varied across time and states allowing the authors to apply a difference-in-difference framework. Using the two data sets Micro Census and Qualification and Career Survey (QaC), PW establish the result of zero returns to schooling. That is, while OLS estimations yield returns in the order of 6-7%, this parameter drops to a number not significantly different from 0% using the IV approach. This result is remarkable as it contradicts most of the evidence from other countries.<sup>3</sup>

Another much acknowledged instrument for education is schooling infrastructure: Card (1995) breaks new ground by exploiting the regional and temporal variation in college proximity. The idea is that the cost of attending a college rises with distance, making the geographical closeness a sufficiently strong indicator of college education. Card uses data from the National Longitudinal Survey (NLS) and finds relatively high IV returns of 13.2% compared to 7.3% OLS returns. Subsequent studies that use the same instrument are Kane and Rouse (1995), Conneely and Uusitalo (1998), and Kling (2001). In a similar spirit, Becker and Siebern-Thomas (2007) (BST henceforth) calculate the returns to education for Germany. Based on SOEP data they use the urbanization of the place of childhood as a proxy for the availability of higher secondary schools. They report returns of about 13% that lie well above their OLS estimates of 6.6%. A third instrument is employed by Ichino and Winter-Ebmer (2004) (IWE henceforth). They use father's involvement in World War II and own educational disruptions due to the war period to instrument education in Germany. Using SOEP data they calculate IV coefficients of 11.3% (own involvement) and 9.4% (father's involvement) for the

---

<sup>3</sup>PW's favorite explanation for this fact is that the basic skills needed for the labor market are learned earlier in Germany than in other countries. An alternative explanation is that the signaling of school track choice, vocational training or apprenticeships is far more important than the actual number of years in schooling.

subpopulations that are affected by the respective instrument.

Table I  
SELECTED IV STUDIES FOR GERMANY

	Data (wave)	Birth cohorts	Instrument	OLS	IV
Piscke & von Wachter (2008)	Micro Census (1989,91,93,95–2004), QaC (1979,85,91,98)	1930–60	compulsory school reform	6–7%	0%
Becker & Siebern-Thomas (2007)	SOEP (1985)	1930–65	urbanization place of childhood	6.6%	13%
Ichino & Winter-Ebmer (2004)	SOEP (1984-86)	1925–49	own ... father's ... World War II involvement	7.6% 7.2%	11.3% 9.4%

Table I reports the main findings of the three German studies. While the OLS coefficients are quite similar across studies, the IV coefficients differ considerably, even hinting at different signs of endogeneity bias. It is the objective of my paper to shed light on these seemingly conflicting results. I claim that the rates of return to education in Germany differ across heterogeneous individuals and that each study correctly identifies the LATE for its respective subsample of compliers: In PW's study this subsample of compliers consists of those who receive one more year of schooling due to the increment in compulsory minimal schooling. Arguably, these were pupils of the basic school track. Pupils of higher tracks did not receive an additional year of schooling due to the reform, as they would have attained more than minimal schooling anyway. For them the IV estimate does not hold. Similar considerations apply for the other two IV studies: BST argue that their instrument provides an interpretation only for schooling differences in grade 10 and above. This is because their instrument, the degree of urbanization of the place of childhood, only affects students of higher tracks because schools of the higher tracks were more likely available in urban places while basic track school were available everywhere. IWE do not document at which grade their instrument affects schooling. Since it is reasonable that World War II involvement did not affect certain educational groups exclusively, their results are likely to apply to a broader subpopulation.

## 4 Estimation approach

### 4.1 Econometric model

Interest lies in parameter  $\delta$  of the linear wage equation

$$W_i = X_i\beta + \delta S_i + u_i \quad (1)$$

where  $W_i$  stands for hourly log wages,  $S_i$  for years of education and  $X_i$  for a  $1 \times k$  vector of exogenous regressors for individual  $i$ . Identification difficulties arise as  $\hat{\delta}_{OLS}$  likely suffers from an endogeneity bias caused by omitted variables, like unobserved ability. To illustrate this paper's estimation strategy that produces estimates free of bias, it is useful to reformulate the endogeneity problem in a control function setting. Rewrite the model as a system of equations

$$W_i = X_i\beta + \delta S_i + u_i \quad (2)$$

$$S_i = X_i\varphi + v_i. \quad (3)$$

$X_i$  may be identical for both equations. We refer to (2) as the wage or primary equation and to (3) as the education or secondary equation. Endogeneity is present if and only if  $cov(u_i, v_i) \neq 0$ . This can be represented by  $u_i = \lambda v_i + e_i$ , where  $v_i$  and  $e_i$  are uncorrelated. Note that one can use this linear combination to replace  $u_i$  in (2) in order to derive the controlled function

$$W = X\beta + \delta S + \lambda v + e, \quad (4)$$

in which  $v$  is called a *control term* (for notational convenience subscripts are dropped). Its impact  $\lambda$  represents the degree of endogeneity in the system. Since  $cov(v, e) = 0$  the controlled equation is free of endogeneity. However, from (3) we know that  $v$  is a perfect linear combination of  $S$  and  $X$ . The regressors are collinear and OLS is infeasible. A standard way to solve this problem is to consider additional regressors  $z$  that restore the orthogonality of  $S$ ,  $X$  and  $v$ . If we could estimate  $v$  with an additional instrument  $z$  so that (3) becomes  $S = X\varphi + z + v$  while  $z$  is not part of (2), the collinearity problem would be solved. The resulting estimates are equivalent to IV.

Rather than pursuing an IV approach, however, this paper follows the approach by Klein and Vella (2010) (KV henceforth), who draw upon second moment exclusion restrictions in the form of heteroskedasticity in order to ensure identification. The key idea of this approach is the notion that one is able to identify the coefficients of interest if the impact of  $v$  is not constant across  $X$  but variable, and that it is possible to estimate this impact. First, replace the unknown  $v$  by its empirical version  $\hat{v}$  that we derive as the residual of (3). Then, transform the control term in (4) as detailed in appendix A.1 to derive the final estimation equation

$$W = X\beta + \delta S + \rho \frac{H_u(X_u)}{H_v(X_v)} \hat{v} + \varepsilon. \quad (5)$$

The parameter  $\rho$  is a correlation coefficient between  $u$  and  $v$  and  $H_u(X_u)$  and  $H_v(X_v)$  represent heteroskedasticity functions of the errors conditional on  $X_u \subseteq X$  and  $X_v \subseteq X$ , respectively.  $X_u$  and  $X_v$  may be different or identical in both equations. Assuming now that the enhanced control term  $(H_u(X_u)/H_v(X_v))\hat{v}$  is not constant across  $X$ , the regressors are no longer collinear and their parameters can be estimated with OLS. Hence, the identifying condition is that  $H_u(X_u)/H_v(X_v) \neq \text{const}$  across  $X$ . That is, identification relies on non-linearity of the control term. KV (2010) call this the *variable impact property* (VIP). A second condition for identification requires that the errors correlation be independent of the regressors and constant, that is  $\text{corr}(uv|X) = \text{corr}(uv)$  and  $\rho = \text{const}$ . This is what the authors call the *constant correlation condition* (CCC).

An argument in favor of the VIP is easy to derive in my application. It simply requires that there is heteroskedasticity in either one or in both equations in a fashion that the quotient of the two functions is non-constant across  $X$ . Various variables of the SOEP dataset like age, length of tenure or sector of employment are potential candidates of causing heteroskedasticity in the wage equation, but not in the education equation. In the education equation, corresponding candidates are number of siblings and ruralness of where one grows up. I provide evidence on this preposition in section 6. Also of little concern is the CCC. It calls for a degree of endogeneity that is constant across regressors. As shown in appendix A.1, all one has to assume is that the correlation between the homoskedastic (unscaled) errors  $\rho = \text{cov}(u^*v^*)$ , i.e. the degree of endogeneity in the model, is

independent of  $X$ . This assumption is essentially (but often concealed) inherent to any IV strategy. Given that both the VIP and the CCC hold, KV prove that identification is established and equation (5) can consistently be estimated. The authors propose a procedure that involves nonparametric estimation of the heteroskedasticity functions  $H_u(X_u)$  and  $H_v(X_v)$ , i.e. without posing any structure on them. The predicted versions of  $\hat{H}_u(X_u)$  and  $\hat{H}_v(X_v)$  are then plugged into (5) and OLS is used to derive the final parameter of  $S$  free of endogeneity bias.

To this day, only a few applications exist that make use of identification through conditional heteroskedasticity as outlined above. Klein and Vella (2009) estimate the return to endogenous schooling decisions for a sample of Australian workers. Farré, Klein, and Vella (2010) perform a similar assessment on a sample of young adults from the U.S. Longitudinal Survey of Youth 1979. Different from Klein and Vella (2009), they show how the estimator can be implemented parametrically, using a non-linear least squares estimator for the heteroskedasticity functions. While this drastically reduces computational costs, it comes at the risk of efficiency losses for miss-specified functional forms. Another exercise of the method comes from Farré, Klein, and Vella (2009), who assess the intergenerational mobility of education. Finally, Schroeder (2010) uses conditional second moments to estimate the impact of microcredit borrowing on household consumption in Bangladesh. I contribute to this small literature by using the same approach to assess the returns to education in Germany.

## 4.2 Implementation

A noticeable advantage of employing nonparametric regression for estimating the heteroskedasticity functions is that it requires neither homoskedasticity nor normality of the errors for consistency. However, their estimation precision decreases dramatically with the inclusion of many regressors, a caveat often referred to as the *curse of dimensionality*. Moreover large sample sizes increase computational costs significantly. For this reason, KV provide proof of consistency also for a semiparametric method, namely Ichimura's (1993) semiparametric least squares (SLS) estimator. SLS reduces the dimension of estimation to a single index and restores the computational feasibility in the presence of many regressors and big

samples.

I follow this approach and employ Ichimura's (1993) SLS in order to estimate the unknown functions  $H_u(X_u)$  and  $H_v(X_v)$ . The estimation of the parameters involves a three step procedure (see Appendix A.2 for a summary of the estimation): In a first step, we use the secondary equation (3) to regress  $S$  on  $X$  in order to obtain consistent estimates of the residuals  $\hat{v}$ . In a second step, we use this residual in squared form to estimate  $\hat{H}_v(X_v)$  with SLS. That is, we derive a parameter  $\hat{\pi}$  as the solution to the minimization problem  $\hat{\pi} = \arg \min_{\pi} \sum \left( \hat{v}^2 - \hat{E}[\hat{v}^2 | X_v \pi_{SLS}] \right)$ . The operator  $\hat{E}[\cdot]$  is the nonparametric conditional expectation.  $X_v \pi$  is a single index that is estimated parametrically as a linear function of  $X_v$ . Index estimators are identified only up to location and scale, that is, they cannot identify a constant and they require that the coefficient of one continuous variable be normalized to one. The predicted  $\hat{v}^2$  can then be used to calculate values for the secondary heteroskedasticity function  $\hat{H}_v(X_v) = \hat{E}[\hat{v}^2 | X_v \hat{\pi}]$ . In a third step, we perform a similar exercise like in step two, but for the primary equation. To be precise, we use  $\hat{v}$  and  $\hat{H}_v(X_v)$  to estimate the missing parameters  $\Phi = [\beta, \delta, \rho]$  and the unknown heteroskedasticity function  $\hat{H}_u(X_u)$  of equation (5) through an iterative procedure: For a given value of  $\Phi$ , say  $\Phi^*$ , that we derive from a naïve OLS estimation of (5), we define the residual to be  $\hat{u}(\Phi^*)$ . Using these residuals regress  $\hat{u}(\Phi^*)^2$  on the single index  $X_u \theta$ , where the parameter  $\theta$  is the SLS estimator of the minimization problem  $\hat{\theta} = \arg \min_{\theta} \sum \left( \hat{u}^2 - \hat{E}[\hat{u}(\Phi^*)^2 | X_u \theta_{SLS}] \right)$ . Now we construct  $\hat{H}_u(X_u) = \hat{E}[\hat{u}(\Phi^*)^2 | X_u \hat{\theta}]$ , just as we have constructed  $\hat{H}_v(X_v)$  in step two, and plug it back into (5). The final estimates of  $\Phi$  and  $\theta_u$  are obtained through an iterative procedure of this last step until convergence of all parameters is achieved. The standard errors are obtained by bootstrapping across all steps. Appendix A.3 provides some comments on the practical aspects of the implementation.

## 5 Data

My analysis draws on the data from the German Socio-Economic Panel Study (SOEP). The SOEP is an annual longitudinal representative household survey. It entails detailed information on income, labor market status, education and a big

range of other socio-economic characteristics. Besides information on the month preceding the interview, it also contains retrospective data on a person's biography. For my analysis the SOEP has two key advantages that makes it preferable over other data sets: First, the SOEP contains a wide variety of socio-economic and biographical variables, which are indispensable for consistently estimating the wage and education equations as well as the two heteroskedasticity functions. As will become apparent from the results in section 6, the rule of thumb is that richer specifications yield more precise estimates. Conversely, parsimoniously specifying any of the equations is detrimental to efficiency. A second argument in favor of the SOEP data is its panel structure, which allows for calculating time-averaged wage incomes within units, thus alleviating potential inefficiencies due to measurement error (Solon 1992).<sup>4</sup> For the purpose of analysis, I use all available annual waves from 1984-2009. The focus rests on full and part time employed workers, excluding self-employed. To reduce censoring of ongoing education activities, I employ information only from respondents who are at least 30 years old and who can safely be assumed to have completed their education. Individuals older than 65 are also excluded because 65 is the legal retirement age for most employees in Germany. The sample is further restricted to persons born after 1939 to exclude potential World War II influences to educational attainment after the age of 6.<sup>5</sup> Moreover, this guarantees that all individuals in the sample received their secondary schooling after the foundation of the Federal Republic of Germany in 1949. East Germany is omitted from the analysis due to the different organization of its educational system during the GDR era. Likewise, the special status of West Berlin during that time makes me exclude this region from the analysis, too.

Table II provides an overview of all variables used in this study. The earnings variable LWAGE3 reports the logarithm of hourly wages. Its information comes from self-reported monthly gross wage income. In order to derive at a measure of hourly wages, the income is divided by the numbers of weekly hours agreed upon in the work contract, times the average number of weeks in a month. To mitigate potential random measurement error that is typically inherent to self-

---

<sup>4</sup>See section 7 for a more thorough discussion of potential measurement error in variables.

<sup>5</sup>Ichino and Winter-Ebmer (2004) report that the cohorts who were in schooling age during World War II have significantly lower educational attainment than other cohorts.

reported information, I calculate a three years moving average measure.<sup>6</sup> Three observations in consecutive years constitute one averaged observation for the year in the middle. Hence, only those observations that have a precedent and a subsequent non-missing value enter the estimation. To approximate the educational attainment I do not use the conventional concept of calculating years of schooling from the highest school degree obtained as mentioned above. Instead, I augment this measure by additional information on the length of a person’s post schooling degree such as degrees from vocational training, apprenticeships, or universities. As a result I arrive at the measure *total years of education* (YRSEDUC), which contains more variation compared to *years of schooling* and, arguably, reflects better the important role of the apprenticeship system and other post-secondary training in the German labor market. This education measure slightly differs from that of PW, but it follows the lines of other studies for the German labor market (e.g. Gang and Zimmermann 2000).

While most variables in Table II are self-explanatory, brief comments on STATE, MIGBACK, and AGEIMMIG are warranted: The STATE of residency dummies comprise a set of 9 binary variables indicating German federal states. For the education equation, I assume that the individuals received most of their education in the state in which they currently live.<sup>7</sup> MIGBACK reports a migration background for all persons of non-German nationality who immigrated to Germany themselves and for those who are of migrant origin but born in Germany. Age at immigration (AGEIMMIG) has a positive integer value for all individuals who have a migration background and immigrated to Germany themselves. For all native Germans and descendants of immigrants born in Germany this variable is coded zero. This way, some individuals have a MIGBACK value of one but an AGEIMMIG value of zero.

The claim of this paper is that the cited IV studies by IWE, PW and BST have a LATE interpretation bounded to the respective subsample of compliers. For IWE they are individuals with all levels of schooling, for PW the compliers are individuals with no or basic school education, and for BST they are individuals with higher school education. In order to substantiate this claim I run

---

<sup>6</sup>An assessment of potential measurement error in section 7 shows the usefulness of this approach.

<sup>7</sup>Pischke (2007) and Siedler (2010) both report that about 85% of the Germans still live in their state of birth, which makes this claim reasonable.



Table II  
VARIABLES DESCRIPTION

LWAGE3	three-years moving average of log hourly gross wage in euros (deflated to 2005)
YRSEDUC	length of secondary and post-secondary education in decimal years
FEMALE	female (yes/no)
BIRTH	year of birth in decimal form (e.g. 1959 = 5.9)
AGE	age at interview
TENURE	length of time with same employer in decimal years
STATE	9 federal state dummies for residency
INDUSTRY	10 NACE classification dummies for sector of employment
FIRM SIZE	4 dummies for number of employees in firm
FULLTIME	working 35 hours weekly or more (yes/no)
EXPER	length of full-time employment experience in decimal years
UNEMPL	length of unemployment experience in decimal years
PUBLIC	employed in the public sector (yes/no)
MIGBACK	migration background (yes/no)
AGEIMMIG	age at immigration (= 0 if born in Germany)
SIBLINGS	number of siblings including half brothers and sisters
RURAL	raised in the countryside (yes/no)

separate regressions for the entire sample (A), for the subsample of individuals holding no school degree or a degree from the lower track (*Hauptschule*) (B), and for the subsample of individuals holding a degree from one of the higher tracks (*Realschule, Fachgymnasium, Gymnasium*) (C). This information is taken from a SOEP variable that reports the highest school degree attained. I use pooled panel data to employ all information available in the data set and increase the number of observations. The sample consists of all person-year observations that have non-missing values for all variables. This amounts to a total of 21,558 observations from 6,066 individuals. The subsample of basic track graduates carries 9,931 observations (2,638 individuals) and the subsample of higher tracks graduates 11,627 (3,584 individuals). On average, each person contributes to the sample with 3.6 observations. Table III displays sample summary statistics.

Klein and Vella's (2010) econometric approach works for an identical set of control variables in both the wage and the education equation. In practice, however, efficiency gains can be exploited by differently specifying both equations according to the nature of their subject. It is self-evident that, for example, job TENURE needs to be included in the wage equation while it does not have any explanatory power in the education equation. The opposite is true, say, for the variable SIB-

Table III  
SAMPLE SUMMARY STATISTICS

	(A)		(B)		(C)	
	ALL		BASIC TRACK		HIGHER TRACKS	
	mean	sd	mean	sd	mean	sd
hourly wage	17.87	(9.02)	14.58	(6.13)	20.22	(10.06)
YRSEDUC	12.30	(2.77)	10.16	(1.00)	13.81	(2.59)
BIRTH	1958	(8.87)	1956	(8.89)	1960	(8.52)
AGE	46.51	(8.23)	47.68	(8.13)	45.49	(8.19)
TENURE	14.41	(10.67)	15.01	(11.01)	13.84	(10.37)
EXPER	19.25	(11.00)	21.82	(11.41)	17.32	(10.21)
UNEMPL	0.42	(1.22)	0.57	(1.52)	0.32	(0.91)
SIBLINGS	2.30	(1.98)	2.85	(2.24)	1.90	(1.66)
FEMALE	0.46		0.41		0.49	
RURAL	0.39		0.44		0.34	
MIGBACK	0.16		0.26		0.10	
PUBLIC	0.30		0.20		0.38	
basic track	0.42		100.00		0.00	
n	6,066		2,638		3,584	
avail. person-years	21,558		9,931		11,627	

Note: SOEP, 1984–2009. Standard errors in parentheses. The sample includes person-year observations for all persons employed in West-Germany (excluding self-employed and West Berlin) born after 1939, and aged between 30 and 65 years, for whom all relevant variables are non-missing. For each individual the information is taken from the latest available year.

LINGS. While the number of siblings seems to be an important piece of information for the education equation,<sup>8</sup> it needs not to be included in the wage equation. Note that regressors that appear in the education equation but not in the wage equation are not necessarily exclusion restrictions in the IV sense that allow for a two stages least squares type of estimator. Besides possessing sufficient explanatory power a valid instrument for education must be uncorrelated with the error terms. A violation of this condition, however, cannot be ruled out per se for any of the variables used in my application. The advantage of KV's method is that one does not have to worry about this problem as identification does not rely on instruments but on the nonlinearity of the control term induced by heteroskedasticity. The

<sup>8</sup>See Becker's (1991) model of the quality-quantity tradeoff of children. It explains the parental choice of the number of children and the human capital investments in each child.

final estimation model takes the following form:

$$\begin{aligned}
\text{LWAGE3} = & \beta_0 + \beta_1 \text{YRSEDUC} + \rho \text{COR} + \beta_2 \text{BIRTH} + \beta_3 \text{AGE} + \\
& \beta_4 \text{AGE}^2 + \beta_5 \text{TENURE} + \beta_6 \text{FEMALE} + \\
& \beta_7 \text{MIGBACK} + \beta_8 \text{PUBLIC} + \beta_9 \text{FULLTIME} + \\
& \beta_{10} \text{EXPER} + \beta_{11} \text{UNEMPL} + \sum_{j=12}^{19} \beta_j \text{STATE} + \\
& \sum_{j=20}^{28} \beta_j \text{INDUSTRY} + \sum_{j=29}^{31} \beta_j \text{FIRMSIZE} + \varepsilon
\end{aligned} \tag{6}$$

$$\begin{aligned}
\text{YRSEDUC} = & \beta_0 + \beta_1 \text{BIRTH} + \beta_2 \text{FEMALE} + \beta_3 \text{SIBLINGS} + \\
& \beta_4 \text{MIGBACK} + \beta_5 \text{AGEIMMIG} + \beta_6 \text{RURAL} + \\
& \sum_{j=7}^{14} \beta_j \text{STATE} + v
\end{aligned} \tag{7}$$

The primary equation (6) is specified in log-linear form. Note that it contains the correction term  $\text{COR} = [H_u(X_u)/H_v(X_v)] \hat{v}$  as an additional regressor and that the variables `SIBLINGS`, `AGEIMMIG`, and `RURAL` appear in the secondary equation but not in the primary equation. Note further that for both equations the regressions of the conditional mean and the heteroskedasticity index are based on the same set of explanatory variables.

## 6 Results

Following the three-step sequential procedure outlined above and displayed in appendix A.2, the output of my approach can be reported in four different estimation outputs: (1) The OLS estimation of the education equation; (2) the SLS estimation of its heteroskedasticity index with  $\hat{v}^2$  as the dependent variable; (3) the OLS estimation of the wage equation; and (4) the SLS estimation of its heteroskedasticity index with  $\hat{u}^2$  as the dependent variable. Note that for the wage equation the results are reported only from the final iteration after convergence was achieved. In order to test my hypothesis that there are heterogeneous returns to education

across graduates from different school types, I perform the entire estimation procedure on the whole sample (A), the subsample of graduates from the basic track (B), and the subsample of graduates from higher school tracks (C). All output tables display the results for the three samples separately.

Table IV  
OLS ESTIMATES — EDUCATION EQUATION

Dependent variable: YRSEDUC			
	(A)	(B)	(C)
	ALL	BASIC TRACK	HIGHER TRACKS
	$\beta_{OLS}$	$\beta_{OLS}$	$\beta_{OLS}$
BIRTH	0.371 *** (0.044)	0.086 *** (0.023)	-0.182 *** (0.064)
FEMALE	-0.175 ** (0.077)	-0.244 *** (0.042)	-0.503 *** (0.103)
RURAL	-0.424 *** (0.079)	0.001 (0.042)	-0.401 *** (0.108)
SIBLINGS	-0.246 *** (0.018)	-0.076 *** (0.010)	-0.138 *** (0.030)
MIGBACK	-2.163 *** (0.231)	-0.923 *** (0.142)	-0.216 (0.362)
AGEIMMIG	0.031 *** (0.009)	0.006 (0.006)	-0.015 (0.012)
c	11.208 *** (0.285)	10.204 *** (0.135)	15.631 *** (0.408)
n	21, 558	9, 931	11, 627
persons	6, 066	2, 638	3, 584
adj. $R^2$	0.14	0.19	0.04

Note: SOEP, 1984–2009. Cluster robust standard errors in parentheses. The sample includes person-year observations for all persons employed in West-Germany (excluding self-employed and West Berlin) born after 1939, and aged between 30 and 65 years, for whom all relevant variables are non-missing. All regressions include STATE dummies. \*= $p < 0.10$ , \*\*= $p < 0.05$ , \*\*\*= $p < 0.01$ .

To reflect the sequence of estimation steps, I begin with presenting the OLS estimates of the education equation in Table IV. The dependent variable is YRSEDUC. Cluster robust standard errors are shown in parentheses. All regressions

include STATE dummies. The adjusted  $R^2$  measures show that the selected socio-economic variables possess a reasonable amount of explanatory power. Only the higher tracks sample has a weak overall goodness-of-fit of 4.1%. Three of the variables are highly significant predictors of education across all samples: Being FEMALE negatively pays off in terms of education. Having many SIBLINGS reduces education, too. The time trend variable birth is positive for samples (A) and (C); it turns negative for sample (B). The remaining variables have different effects across the three samples: Children who grew up in RURAL areas are educationally disadvantaged, except for the basic track sample where the coefficient is insignificant. A possible explanation is that having a rural background is itself an indicator for the track choice as also visible in the summary statistics of Table III. A similar reasoning might help to understand the coefficients of MIGBACK. For the whole sample the effect is negative and significant. For higher tracks, however, this association vanishes. Again we observe from Table III strong self-selection of migrants into the lower school track. The AGEIMMIG coefficient is positive but small for the whole sample. It renders insignificant in the two subsamples.

The results of the SLS heteroskedasticity index estimation are reported in Table A-I of the appendix. The coefficient normalized to one is BIRTH. Given this normalization, coefficients can only be interpreted in relative terms. The dependent variable is the squared residual from the education equation,  $\hat{v}^2$ . As the coefficients of the index do not have an immediate interpretation they are not discussed. Note, however, that some are statistically significant under the assumption of normality<sup>9</sup> and that the overall goodness-of-fit, measured by  $R^2$ , is considerable for all three samples. This suggests that heteroskedasticity is present and that the VIP is likely to hold. Further evidence for the presence of heteroskedasticity comes from formal tests, which I only report for the whole sample: Table A-II in the appendix shows the results from regressing the squared residuals from the education equation onto all explanatory variables. The joint  $\chi^2$  test statistic rejects the null hypothesis of homoskedasticity at any common significance level. Graphical analyses in Figure A-I substantiate the claim of heteroskedasticity being at work, too. For the whole sample, panels A and B show that the variance of the residuals decreases in both

---

<sup>9</sup>This assumption becomes reasonable as the sample size increases, because the SLS estimator is asymptotically normally distributed (Ichimura 1993).

the number of SIBLINGS and the year of BIRTH. Similarly, panel C suggests that having a migration background decreases the variance. The dispersion in education levels is different between natives and immigrants. For the variable RURAL in panel D there is no clear relationship visible although its parameter in the SLS estimation and its heteroskedasticity test are strongly significant.

Table V displays the main results for the wage equation. For each sample two outputs are presented: one from a naïve OLS regression ignoring the endogeneity of schooling, and one of the last iteration from the control function approach with the control term  $COR = [H_u(X_u)/H_v(X_v)] \hat{v}$  as an additional regressor. In both the OLS and the CF settings the wage equation is specified as a linear function. Its semi-log characteristic allows for interpreting the coefficients as percentage changes in wage income given a unit change in the regressors. Except for the control term coefficient  $\rho$ , the interpretation of the CF parameter estimates is identical to OLS.

The t-statistics for the OLS estimates are corrected to be robust against repeated observations for the same individual. The CF standard errors come from bootstrapping across the whole procedure with 400 replications. The p-values are derived by means of the percentile method. In order to account for the clustered data, sampling is realized across cluster units with replacement. All regressions include controls for STATE, INDUSTRY, and FIRMSIZE. The sign and size of the control term coefficient,  $\rho$ , allows direct inference on the direction and size of the endogeneity in the model. A positive coefficient indicates that OLS parameters of YRSEDUC are upward biased. A negative coefficient reports a downward bias. The level of statistical significance informs about the extent to which the model successfully employs the VIP for identification. A significant correction term indicates both that there is endogeneity and that the estimator performs well in purging it from the model. Conversely, an insignificant correction term hints at identification failure, which goes back to either poor performance of the estimator or a lack of endogeneity. While a detailed examination of each estimation step may reveal insights to the performance of the estimator and theory may guide considerations about the existence of endogeneity, one cannot test formally which one of the two effects is at work.

The central result of my paper becomes apparent from the estimates of the key parameters  $\beta_1$  and  $\rho$ , which belong to YRSEDUC and COR, respectively. The results

Table V  
OLS & CF ESTIMATES — WAGE EQUATION

Dependent variable: LWAGE3											
	(A)		(B)				(C)				
	ALL		BASIC SCHOOL TRACK				HIGHER SCHOOL TRACKS				
	$\beta_{OLS}$	$\beta_{CF}$	$\beta_{OLS}$	$\beta_{CF}$	$\beta_{OLS}$	$\beta_{CF}$	$\beta_{OLS}$	$\beta_{CF}$	$\beta_{OLS}$	$\beta_{CF}$	
YRSEDUC	0.073 *** (0.002)	0.085 *** (0.006)	0.0442 *** (0.005)	0.023 *** (0.009)	0.0706 *** (0.002)	0.110 *** (0.011)					
COR		-0.116 * (0.057)		0.114 *** (0.037)		-0.361 ** (0.099)					
BIRTH	0.042 *** (0.006)	0.038 *** (0.006)	0.065 *** (0.008)	0.068 *** (0.008)	0.015 * (0.009)	0.022 ** (0.008)					
AGE	0.023 *** (0.004)	0.023 *** (0.004)	0.015 *** (0.005)	0.015 *** (0.005)	0.028 *** (0.005)	0.029 *** (0.006)					
AGE <sup>2</sup>	0.000 *** (0.000)	0.000 *** (0.000)	0.000 ** (0.000)	0.000 ** (0.000)	0.000 *** (0.000)	0.000 *** (0.000)					
TENURE	0.005 *** (0.000)	0.005 *** (0.000)	0.004 *** (0.001)	0.004 *** (0.001)	0.005 *** (0.001)	0.006 *** (0.001)					
FEMALE	-0.170 *** (0.009)	-0.169 *** (0.009)	-0.214 *** (0.014)	-0.222 *** (0.015)	-0.143 *** (0.012)	-0.126 *** (0.014)					
MIGBACK	-0.090 *** (0.010)	-0.069 *** (0.015)	-0.071 *** (0.012)	-0.080 *** (0.013)	-0.145 *** (0.017)	-0.117 *** (0.018)					
PUBLIC	-0.026 *** (0.009)	-0.027 *** (0.009)	0.011 (0.015)	0.013 (0.015)	-0.048 *** (0.012)	-0.047 *** (0.012)					
FULLTIME	0.036 *** (0.011)	0.035 *** (0.012)	0.058 *** (0.017)	0.060 *** (0.018)	0.029 * (0.015)	0.028 ** (0.015)					
EXPER	0.004 *** (0.001)	0.004 *** (0.001)	0.003 *** (0.001)	0.003 *** (0.001)	0.004 *** (0.001)	0.005 *** (0.001)					
UNEMPL	-0.024 *** (0.003)	-0.023 *** (0.003)	-0.020 *** (0.004)	-0.020 *** (0.004)	-0.036 *** (0.006)	-0.035 *** (0.006)					
c	0.393 *** (0.101)	0.263 *** (0.112)	0.725 *** (0.140)	0.926 *** (0.147)	0.507 *** (0.141)	-0.108 (0.199)					
n	21,558	21,558	9,931	9,931	11,627	11,627					
persons	6,066	6,066	2,638	2,638	3,584	3,584					
adj. R <sup>2</sup>	0.51	0.52	0.42	0.42	0.49	0.49					

Note: SOEP, 1984–2009. Cluster robust standard errors in parentheses. The sample includes person-year observations for all persons employed in West-Germany (excluding self-employed and West Berlin) born after 1939, and aged between 30 and 65 years, for whom all relevant variables are non-missing. The CF standard errors and p-statistics are calculated from clustered bootstrapping with 400 replications. All specifications of the wage equation additionally control for STATE dummies, INDUSTRY sector dummies and FIRM SIZE dummies. The wage heteroskedasticity function is estimated via SLS using the control variables presented in Table A-III. The school equation is estimated via OLS, its heteroskedasticity function with SLS. Both control for the same set of variables as presented in Table IV. \*= $p < 0.10$ , \*\*= $p < 0.05$ , \*\*\*= $p < 0.01$ .

are presented in the first two rows of Table V. The OLS estimation for the sample (A) yields a 7.3% return from one additional year of education. In the controlled

setting, this parameter increases by only 1.2 percentage points to 8.5%, hinting at only a small downward bias in the naïve model. The corresponding coefficient of the correction term,  $\rho$ , is only weakly significant at the 10% level. The picture changes for the two subsamples: Both start off with positive returns to education from OLS estimations, with 4.4% for the basic track graduates and 7.1% for the higher track graduates. When applying the CF estimation procedure, however, the results diverge considerably. For the lower educated individuals of sample (B) the control term parameter is positive and strongly significant. It shrinks the education coefficient to 2.3% while retaining its precision. That is, after accounting for the endogeneity in the model, there remains a slightly positive return from one additional year of education for graduates from the basic school track. Opposite results are found with better educated individuals in sample (C). Here, the control term parameter is negative and significant. It pushes the OLS estimates from 7.1% to 11%, also retaining its precision. Altogether, the diverging results of the two subsamples suggest that there are heterogeneous returns to education across the population: Not being bounded by LATE interpretations, my causal CF estimations with identification via heteroskedasticity reveal that additional education pays off generously for graduates from a higher school track, while there is little effect for graduates from the basic school track and individuals with no secondary schooling.

The remaining parameter estimates of the control variables display a rather homogenous pattern across the three samples. In line with most wage regressions, AGE positively affects wages. However, different than expected AGE<sup>2</sup> indicates that it does so in a linear fashion. The variables BIRTH cohort, job TENURE, FULLTIME employment and job market experience (EXPER) also increase hourly wages. Instead, being FEMALE, having a migration background (MIGBACK) and possessing increasing unemployment experience (UNEMPL) depress wages. For the basic track subsample the parameter for PUBLIC is insignificant. This may stem from the fact that in the basic track sample only a share of 20% is employed in the public sector (cf. Table III).

The results of the wage equation's heteroskedasticity estimation are reported in Table A-III in the appendix. Note that they come from the last iteration after achieving convergence. The parameter set to one is YRSEDUC. As with the edu-



cation equation, the coefficients of the SLS regression do not have a meaningful interpretation. Note that some parameters are statistically significant and that the overall goodness-of-fit for all three samples is reasonably high. This indicates the presence of heteroskedasticity. This proposition is further supported by formal tests in Table A-IV, which again are only presented for the whole sample.<sup>10</sup> The joint  $\chi^2$  test statistic strongly rejects the null hypothesis of homoskedasticity. Likely candidates for heteroskedasticity are YRSEDUC, AGE, TENURE and MIGBACK among others. The graphical analyses in Figure A-II depict exemplary how these variables might influence the squared residual  $\hat{u}^2$ . Visual inspection of panel A reveals that between YRSEDUC and the error variance there seems to be a nonlinear relationship with different slopes across the variable's range. AGE in panel B does not exhibit a visible pattern although the chi-squared test indicates a significant influence on  $\hat{u}^2$ . The scatter plot for TENURE in panel C reveals that working long with the same employer reduces the variability of the error term. Finally, being immigrant (MIGBACK) in panel D seems to reduce it, too.

Reviewing figures A-I and A-II suggests that the ATE interpretation of my estimates is justified. My parameters are ATEs only if the VIP holds for the whole sample. In other words, there must be heteroskedasticity across the whole range of at least one variable in at least one equation, and if present in both equations, the conditional variance functions must differ from one another. Visual inspection of the scatterplots suggests that the heteroskedasticity is not restricted to a certain range of the regressors' values. Moreover, identical variance functions in the wage and education equations are very unlikely. The simple fact that SIBLINGS and RURAL cause heteroskedasticity in the education equation, but are not part of the wage equation, supports this notion. Variables shared by both equations also reveal a diverging pattern across equations, as the variable MIGBACK shows exemplarily. Final support for the validity of the VIP comes from the fact that for the two subsamples the control term parameter is well identified. This shows that the heteroskedasticity functions of the education equation and the wage equation are not identical and that the VIP holds.

The OLS and CF results in Table V confirm the IV literature on returns to education: Starting with the benchmark OLS results for the whole sample, my 7.3%

---

<sup>10</sup>The test statistics for the two subsamples are available from the author upon request.

returns lie well in the range of 6-7.3% defined by the above discussed literature (cf. Table I). For the two subsamples, I do not draw comparisons with respect to OLS results because most IV studies do not split the sample into educational subgroups. Moving on to the more interesting CF estimates, recall the argument that my three different samples approximate the LATE subgroups of Ichino and Winter-Ebmer (2004), Pischke and Wachter (2008), and Becker and Siebern-Thomas (2007), respectively. Beginning with IWE, one can view the treatments of their instruments, own and father's World War II involvement, to be not restricted to one particular subgroup of the population. Given this is true, their LATE represents closely the true ATE. Their two different instruments yield returns of 9.4% and 11.3%, respectively. My CF estimate for the whole sample of 8.5% is smaller than these numbers, yet not very far away from them. Moving on to PW, I draw upon the subsample of the lower educated. Just like the authors I find strong evidence for upward biased OLS estimates that diminish dramatically once the endogeneity is controlled for. My causal return to education is 2.3%. Although not 0%, this number is in line with PW's "zero returns to schooling" regarding the conclusion that the returns for lower educated persons are considerably lower than naïve OLS suggests. Finally, BST calculate a causal return of 13%. They acknowledge that this result can, most likely, only be interpreted for graduates from grade 10 and above. The comparison group in my study is, therefore, the individuals who graduated from higher school tracks. Just like in BST my calculations indicate downward biased OLS estimates that are raised to 11% when controlling for the endogeneity. While this finding contrasts with the interpretation of omitted ability bias, it matches well BST's result and also reflects much of the international IV literature (cf. section 3). I conclude that PW's 0% return to schooling and BST's 13% return to schooling are just two sides of the same coin. Both findings just reflect different parameters for different subsamples of the population. The study that most likely finds the population's ATE is the one by IWE with a range of return of 9.4-11.3%.

Before concluding this section, it is worth noting that, at first glance, my results appear to contradict the common hypothesis of diminishing returns to education, *i.e.* the assumed concave relationship between education and labor market income (Becker 1993). The hypothesis predicts that individuals with few years of educa-

tion receive a higher wage premium for one additional year than individuals with many years of education. In contrast, I find a low wage premium for lower educated and a high wage premium for higher educated. However, this contradiction is spurious. The identification approach employed in this paper, as introduced in Klein and Vella (2010), essentially builds on the assumption of linearity of the primary equation. This allows identifying only one single slope parameter for the entire sample. Causality is established *within* the sample. Splitting the sample in two still allows for causal inference within each sample. It does not, however, allow for causal inference *across* the samples as self-selection into higher and lower education is still a problem. In other words, while the estimated parameters of 2.3% and 11.0% are average returns for a randomly selected individual *within* the group of lower and higher educated individuals, respectively, they cannot inform about different returns at different points of the educational scale of a randomly selected individual from the *entire* sample. Within the entire sample, the average return is 8.5%.

## 7 Robustness checks

This section presents four different checks to evaluate the robustness of my results: First, I vary the specification of the primary equation in order to assess the estimates' sensitivity to a changing number of control variables. Secondly, I use different samples, which allows resembling the literature's samples even closer. Thirdly, I split the sample into two halves across birth cohorts to check for variations over time. Finally, I assess the choice of the dependent variable by repeating the estimation for log wages with no averaging and log wages with a five-years moving average. This final exercise allows conjectures about the potential problems of measurement error in the variables.

1. Varying the specification: A major difference of my approach to IV studies is that it requires a relatively rich specification of the equations in order to guarantee a maximum degree of precision at every stage of the estimation procedure. Identifying causal parameters via different forms of heteroskedasticity in two equations is only feasible if enough heteroskedasticity generating variables are included in the model and if the heteroskedasticity functions can be estimated at

a reasonable fit. Replicating IV baseline specifications without any controls will not be successful using Klein and Vella's (2010) method. In order to gain some insights into the extent to which my CF results hinge on a rich specification, Table VI reports the estimation results for different sets of control variables while holding the sample size constant. As I opt to employ the same variable matrix for the level regression as for the heteroskedasticity regression, this also affects the SLS estimation of the squared residuals. Hence, one effect I expect is a loss in precision due to a poorer fit in the heteroskedasticity estimation. An opposing effect might occur if the correction term is correlated with left-out control variables in the level estimation. In this case, part of the captured variance may translate into increased significance of  $\rho$ . The overall effect on the precision of parsimonious specifications remains ambiguous. Now turning to changes not in the precision but in the *value* of parameters, the stepwise omission of control variables allows to disentangle the partial effect of education from its total effect. For this reason I successively remove all variables that are potentially correlated with YRSEDUC (PUBLIC, FULLTIME, TENURE, EXPER, UNEMPL, FIRMSIZE, INDUSTRY and STATE) and leave all variables in the model that are considered to be predetermined with respect to education (BIRTH, AGE, AGE<sup>2</sup>, FEMALE, MIGBACK).

Table VI displays the estimation results for the two key variables YRSEDUC and COR for four different specifications (columns 1-4) and the three samples (A), (B) and (C). Column 1 mirrors the results from the preferred specification of equations (6) and (7) that we know already from Table V. This specification employs the full set of available control variables. Moving from left to right the number of controls decreases. Column 2 neither controls for FIRMSIZE, INDUSTRY and STATE dummies in the wage equation nor in the SLS estimation of the heteroskedasticity index. Column 3 additionally disregards the variables TENURE, EXPER and UNEMPL. And column 4 does so for PUBLIC and FULLTIME.

One can see that the key parameters do not change very much: For sample (A), COR is negative and weakly (if at all) significant. The OLS and CF estimates of YRSEDUC differ only slightly. For sample (B), COR is positive and significant across all equations. One can observe, however, that the level of significance slightly decreases as the specification becomes sparse. This is also true for the YRSEDUC estimate. Its value, however, remains closely in the vicinity of 2%. For

Table VI  
VARYING THE SPECIFICATION

	(1)		(2)		(3)		(4)	
	$\beta_{OLS}$	$\beta_{CF}$	$\beta_{OLS}$	$\beta_{CF}$	$\beta_{OLS}$	$\beta_{CF}$	$\beta_{OLS}$	$\beta_{CF}$
(A) ALL SCHOOL TRACKS								
YRSEDUC	0.073 *** (0.002)	0.085 *** (0.006)	0.076 *** (0.002)	0.087 *** (0.005)	0.071 *** (0.002)	0.072 *** (0.005)	0.071 *** (0.002)	0.075 *** (0.004)
COR		-0.116 * (0.057)		-0.101 * (0.049)		-0.012 (0.042)		-0.036 * (0.034)
n	21,558		21,558		21,558		21,558	
(B) BASIC SCHOOL TRACK								
YRSEDUC	0.044 *** (0.005)	0.023 *** (0.009)	0.044 *** (0.006)	0.026 *** (0.009)	0.044 *** (0.006)	0.024 ** (0.009)	0.045 *** (0.006)	0.024 ** (0.009)
COR		0.144 *** (0.037)		0.090 *** (0.035)		0.100 ** (0.034)		0.102 ** (0.031)
n	9,931		9,931		9,931		9,931	
(C) HIGHER SCHOOL TRACKS								
YRSEDUC	0.071 *** (0.002)	0.110 *** (0.011)	0.073 *** (0.002)	0.104 *** (0.012)	0.065 *** (0.002)	0.084 <sup>a</sup> *** (0.009)	0.065 *** (0.002)	0.102 *** (0.009)
COR		-0.361 ** (0.099)		-0.266 * (0.097)		-0.167 <sup>a</sup> (0.073)		-0.308 * (0.071)
n	11,627		11,627		11,627		11,627	
<i>Control variables:</i>								
PUBLIC	yes	yes	yes	yes	yes	yes	no	no
FULL-TIME	yes	yes	yes	yes	yes	yes	no	no
TENURE	yes	yes	yes	yes	no	no	no	no
EXPER	yes	yes	yes	yes	no	no	no	no
UNEMPL	yes	yes	yes	yes	no	no	no	no
FIRMSIZE	yes	yes	no	no	no	no	no	no
INDUSTRY	yes	yes	no	no	no	no	no	no
STATE	yes	yes	no	no	no	no	no	no

Note: SOEP, 1984–2009. Cluster robust standard errors in parentheses. The sample includes person-year observations for all persons employed in West-Germany (excluding self-employed and West Berlin) born after 1939, and aged between 30 and 65 years, for whom all relevant variables are non-missing. The CF standard errors and p-statistics are calculated from clustered bootstrapping with 400 replications. All specifications control for the predetermined variables BIRTH, AGE, AGE<sup>2</sup>, FEMALE and MIGBACK. The wage heteroskedasticity function is estimated via SLS using the control variables presented in Table A-III. The school equation is estimated via OLS, its heteroskedasticity function with SLS. Both control for the same set of variables as presented in Table IV. \*= $p < 0.10$ , \*\*= $p < 0.05$ , \*\*\*= $p < 0.01$ . a) No convergence achieved after 50 iterations (the displayed values come from the 50<sup>th</sup> iteration).

sample (C), the decline in significance of COR is more pronounced. Nonetheless, the YRSEDUC parameter values all remain in the range of 10-11%, except for column 3 where convergence fails. All in all, reducing the number of controls somewhat

affects precision. However, the point estimates themselves change only marginally. Striking differences between the partial and total effect of education cannot be detected.

2. Varying the sample: My sample definition is typical for the returns-to-education literature but does not perfectly match the ones used in the cited literature. PW just like IWE, for example, exclude individuals with migration background from their analyses. Angrist and Krueger (1991) restrict their analysis to men. BST limit their sample to the full-time employed. Such narrow definitions clearly circumvent potential problems of unexplained heterogeneity from the peculiarities of, say, female and immigrant employment or non-linear effects in part-time occupation. Different from such IV studies, however, my approach relies on a rich set of control variables that potentially induce heteroskedastic error terms. Restricting my sample to native full-time employed males in, say, the private sector removes four variables (MIGBACK, FULLTIME, FEMALE, and PUBLIC) from the model, all of which proved to be a source for heteroskedasticity and are therefore important for identification. In order to investigate the robustness of my results not only to different specifications but also to different sample definitions, Table VII reports the OLS and CF estimates for four different samples: Column 1 excludes immigrants from consideration, column 2 restricts the sample to full-time employed individuals, column 3 focuses on males only, and column 4 excludes public sector employees. For each sample definition, one control variable must be dropped from the model. For example, only focusing on natives makes the dummy MIGBACK redundant.

Browsing across columns 1-4 for all three samples (A), (B), and (C) reveals that, by and large, the point estimates are somewhat resistant against varying sample definitions, while their precision is more strongly affected. This feature is most pronounced for sample (A), where COR is significant only in panel 1-A, which most closely resembles the sample definition of IWE. The results are comparable to the baseline specification. Sample (C) reveals a similar pattern to sample (A), although the point estimates fluctuate more. In panel 2-C, which is closest to the sample definition of BST, the 11% causal return to education is identical to the

baseline result.<sup>11</sup>

Table VII  
VARYING THE SAMPLE

	(1)		(2)		(3)		(4)	
	ONLY NATIVES		ONLY FULL-TIME EMPLOYED		ONLY MALES		ONLY PRIVATE SECTOR	
	$\beta_{OLS}$	$\beta_{CF}$	$\beta_{OLS}$	$\beta_{CF}$	$\beta_{OLS}$	$\beta_{CF}$	$\beta_{OLS}$	$\beta_{CF}$
(A) ALL SCHOOL TRACKS								
YRSEDUC	0.077 *** (0.002)	0.088 *** (0.006)	0.072 *** (0.002)	0.083 *** (0.007)	0.069 *** (0.003)	0.085 *** (0.008)	0.071 *** (0.002)	0.089 *** (0.008)
COR		-0.127 ** (0.060)		-0.109 (0.068)		-0.159 (0.077)		-0.152 (0.066)
n	17,593		17,155		12,513		14,892	
(B) BASIC SCHOOL TRACK								
YRSEDUC	0.076 *** (0.008)	0.049 <sup>a</sup> *** (0.014)	0.046 *** (0.006)	0.025 ** (0.011)	0.041 *** (0.007)	0.021 * (0.011)	0.046 *** (0.006)	0.023 ** (0.009)
COR		0.025 <sup>a</sup> (0.047)		0.120 *** (0.045)		0.115 *** (0.043)		0.125 *** (0.042)
n	7,111		8,195		6,159		7,823	
(C) HIGHER SCHOOL TRACKS								
YRSEDUC	0.071 *** (0.002)	0.092 *** (0.009)	0.069 *** (0.003)	0.110 *** (0.012)	0.068 *** (0.004)	0.078 *** (0.014)	0.066 *** (0.003)	0.093 *** (0.011)
COR		-0.207 ** (0.088)		-0.388 (0.109)		-0.102 (0.124)		-0.222 (0.089)
n	10,482		8,960		6,354		7,069	
<i>Control variables:</i>								
MIGBACK	no	no	yes	yes	yes	yes	yes	yes
FULLTIME	yes	yes	no	no	yes	yes	yes	yes
FEMALE	yes	yes	yes	yes	no	no	yes	yes
PUBLIC	yes	yes	yes	yes	yes	yes	no	no

Note: SOEP, 1984–2009. Cluster robust standard errors in parentheses. The sample includes person-year observations for all persons employed in West-Germany (excluding self-employed and West Berlin) born after 1939, and aged between 30 and 65 years, for whom all relevant variables are non-missing. The CF standard errors and p-statistics are calculated from clustered bootstrapping with 400 replications. All specifications control for the predetermined variables BIRTH, AGE, AGE<sup>2</sup>, FEMALE and MIGBACK. Additional controls are STATE dummies, INDUSTRY sector dummies and FIRM SIZE dummies. The wage heteroskedasticity function is estimated via SLS using the control variables presented in Table A-III. The school equation is estimated via OLS, its heteroskedasticity function with SLS. Both control for the same set of variables as presented in Table IV. \*= $p < 0.10$ , \*\*= $p < 0.05$ , \*\*\*= $p < 0.01$ . a) No convergence achieved after 50 iterations (the displayed values come from the 50<sup>th</sup> iteration).

A notable exception from this regularity is panel 1-B which most closely resembles PW's sample definition. Convergence fails and the correction term does not

<sup>11</sup>One can replicate the sample definition by BST even more closely by restricting birth cohorts to 1930-1960. The results (not reported here) are only marginally different.

capture much of the endogeneity. However, in columns 2-4 the estimator behaves similar to the baseline setting and does not reveal a notable influence of the more restrictive sample definitions.

3. Splitting the sample across time: One concern is that the returns to education in Germany changed over time and that my estimations only apply to certain cohorts of the population. In order to assess the sensitivity of my results across time I split each of the educational subsamples (A), (B), and (C) into two groups using the median birth year as the cutoff point. This way, two groups are generated that allow a comparison between early and late born individuals. As the instruments used by PW, BST and IWE predominantly affect early birth cohorts, this is another way of validating the results in the literature. Table VIII displays the median cutoff values, the number of observations in each group and the OLS and CF estimation results.

The comforting feature of this intertemporal comparison is that the results for the early birth cohorts are very similar to the baseline results in terms of level estimations and precision, which corroborates the established link between my results and the existing literature. My main results hold for the cohorts born between 1939 and 1956/59/60. Note that the coefficient for the entire sample of 9.7% now falls into IWE's range of 9.4-11.3%. When moving on to the late birth cohorts, however, the situation is different. OLS returns decrease considerably and so do the causal returns, except for the lower educated in sample (B). While the sign of the endogeneity correction parameter remains intact, its significance vanishes. For sample (B) also the magnitude of the correction term is negatively affected by focusing on the later-born. It remains unclear, however, if the failure to correct for endogeneity for the later birth cohorts results from the fact that the returns to education have decreased over time, or from the fact that the endogeneity in estimating them has decreased, or because there is a violation of the VIP assumption.

4. Assessing potential measurement error: One issue not sufficiently addressed so far is measurement error (ME) in the variables. Both of this study's key variables, hourly wages and YRSEDUC, are self-reported and may hence suffer from imprecisions. The potential consequences for the point estimates differ by variable:

While ME in the dependent variable *hourly wages* adds noise to the estimation,



Table VIII  
COMPARING EARLY AND LATE BIRTH COHORTS

	(1)		(2)	
	$\leq$ MEDIAN		$>$ MEDIAN	
	$\beta_{OLS}$	$\beta_{CF}$	$\beta_{OLS}$	$\beta_{CF}$
(A) ALL SCHOOL TRACKS (median = 1959)				
YRSEDUC	0.078 *** (0.002)	0.097 *** (0.005)	0.067 *** (0.002)	0.078 *** (0.008)
COR		(-0.181) ** (0.048)		-0.104 (0.077)
n	11,354		10,204	
(B) BASIC SCHOOL TRACK (median = 1956)				
YRSEDUC	0.047 *** (0.008)	0.020 ** (0.011)	0.039 *** (0.007)	0.034 ** (0.010)
COR		(0.153) ** (0.048)		0.028 (0.045)
n	5,015		4,916	
(C) HIGHER SCHOOL TRACKS (median = 1960)				
YRSEDUC	0.075 *** (0.003)	0.092 *** (0.007)	0.066 *** (0.004)	0.085 *** (0.012)
COR		-0.173 ** (0.070)		-0.179 (0.108)
n	6,140		5,487	

Note: See Table V.

it does not bias the estimators. For the right-hand side variable YRSEDUC the issue of ME is more serious as it may lead to biased estimates. For panel data, one way to mitigate the impact of random ME is to take averages over time for subsequent observations from the same individual (Solon 1992). This strategy is meaningless for time-*constant* information like years of education, but it does have potential precision gains for time-*varying* variables like wages. In fact, for the wage, this approach is implemented already throughout all analyses by taking three-years moving averages (LWAGE3). A tentative idea of how such averaging helps to mitigate estimation imprecisions due to ME can be derived from Table

IX. Column 2 reproduces the benchmark results with three-years moving average of wages, LWAGE3, as known already from Table V. Column 1 uses plain (not averaged) observations, LWAGE1, and column 3 uses five-years moving averages, LWAGE5. All wages are in logarithms. Moving from left to right, one observes how the number of observations decreases as the data demands increases. LWAGE3 requires three successive non-missing observations increases. LWAGE5 requires five.

The numbers show that the point estimates for all three subsamples are quite robust to changes in the dependent variable. The existing but small differences can be driven by both the choice of the endogenous variable and the change in number of observations. For assessing potential measurement error most interest lies in the estimates' precision. Moving from left to right reveals that the precision remains unchanged for sample (A), it increases for sample (B), and it decreases for sample (C). While this is no strong indication of measurement error in the overall sample, time-averaged measures help to improve the estimation at least for the subsample of the lower educated. Taking LWAGE3 as the dependent variables turns out to be a reasonable, at least not harmful, choice in the trade-off between ME correction and the preservation of a fair number of observations.

Turning to the potential measurement error in the education variable, the matter becomes more complicated. The observed YRSEDUC might differ from true years of education for three reasons: First, the discrete assignment of years to educational attainments might not correctly mirror the actual time spent to achieve them. Consider the example of an individual with a middle school track degree and an apprenticeship. For the school degree a fixed value of 10 years is assigned, regardless of how many years the individual actually spent in school. Similarly, the apprenticeship scores 2 years although some apprenticeships take longer and some take shorter than that. This discrete assignment results in a random or *classical measurement error*. Second, the variable YRSEDUC is censored at both tails. At the lower end, individuals with less than 7 years of schooling are nevertheless assigned 7 years resulting in a positive ME. At the upper end, the maximum score is 18 years. For individuals with more than 18 years of education, this produces a negative ME. Hence, outside the range of allowed years of education the ME is negatively correlated with YRSEDUC resulting in *mean-regressive measurement*

Table IX  
MEASUREMENT ERROR: CHANGING THE DEPENDENT VARIABLE

Dep. var.	(1) LWAGE1		(2) LWAGE3		(1) LWAGE5	
	$\beta_{OLS}$	$\beta_{CF}$	$\beta_{OLS}$	$\beta_{CF}$	$\beta_{OLS}$	$\beta_{CF}$
(A) ALL SCHOOL TRACKS						
YRSEDUC	0.076 *** (0.001)	0.084 *** (0.005)	0.073 *** (0.002)	0.085 *** (0.006)	0.071 *** (0.002)	0.082 *** (0.007)
COR		-0.066 * (0.042)		-0.116 * (0.057)		-0.101 * (0.069)
n	31,289		21,558		14,945	
(B) BASIC SCHOOL TRACK						
YRSEDUC	0.050 *** (0.005)	0.037 *** (0.009)	0.044 *** (0.005)	0.023 *** (0.009)	0.042 *** (0.006)	0.024 *** (0.008)
COR		0.056 * (0.034)		0.114 *** (0.037)		0.108 *** (0.038)
n	14,199		9,931		7,239	
(C) HIGHER SCHOOL TRACKS						
YRSEDUC	0.073 *** (0.002)	0.107 *** (0.010)	0.071 *** (0.002)	0.110 *** (0.011)	0.069 *** (0.003)	0.099 *** (0.010)
COR		-0.254 *** (0.076)		-0.361 ** (0.099)		-0.294 * (0.101)
n	17,090		11,627		7,706	

Note: See note of Table V.

*error*.<sup>12</sup> A third source of ME has very similar characteristics. Questionnaire respondents may simply misreport their educational attainments. At intermediate levels of the variable's range this error may be random. At the lower and upper limits, however, it is more plausible to assume a mean-regressive ME. Specifically, individuals with very low levels of education cannot under-report education, whereas individuals with very high levels of education cannot over-report. Again, the ME is negatively correlated with YRSEDUC and therefore mean-regressive.

IV estimations for non-categorical variables are usually consistent under classical measurement error. IV does not, however, guard against mean-regressive ME. In contrast, my CF approach is vulnerable to both classical and mean-regressive measurement error. Classical ME leads to an attenuation bias in the coefficient towards zero. The magnitude of this attenuation increases upon the inclusion of other independent variables that are correlated with the wrongly measured independent variable (Bound, Brown, and Mathiowetz 2001). The bias of mean-regressive ME

<sup>12</sup>Mean-regressive measurement error has, in fact, received some acknowledgment in the return to education literature e.g. by Kane, Rouse, and Staiger (1999), Bound and Solon (1999), and Black, Berger, and Scott (2000).

is more difficult to determine. In general, however, in absolute terms, the coefficient of the variable with mean-regressive ME lies above the coefficient that would result with only classical ME (Bound, Brown, and Mathiowetz 2001). Whether it also exceeds the true coefficient depends on the strength of the relationship between the true years of education and the ME. As long as the relationship is weak, it does not. This happens for a few bottom and top codings. Then the ME is random for most observations. The share of observations that is neither bottom nor top coded amounts to 87.6% for the whole sample, 94.5% for the basic school track sample, and 77.9% for the higher school track sample. Correspondingly, I suppose that the impact of mean-regressive measurement error varies across the samples. It is highest for the higher track sample and lowest for the basic track sample. The latter has almost no bottom and top codings. Conclusions for my parameter estimates are hardly derivable from these numbers. I know, however, that the attenuation bias in both the OLS and the CF estimates are bigger for the basic track graduates than for higher tracks graduates. Since the share of bottom and top coded observations in neither sample is excessively large, I suspect that true parameters lie slightly above my estimates.

Before concluding, a word of caution about the interpretation of the coefficients in the return to education literature, specifically for studies about Germany, is warranted: The conventional concept of measuring education in years of schooling closely resembles the concept of human capital accumulation. Even so, it is not necessarily convincing for a structured school system like the one in Germany where different tracks lead to different degrees with little mobility from one to another. The problem arises when researchers claim to measure the *quantity* of schooling while holding the *quality* of schooling constant. For Germany this is most likely not the case. One cannot assume that, for example, one year of the basic track (*Hauptschule*) is equivalent to one year of the highest track (*Gymnasium*). Regression results for Germany should therefore not be interpreted from the perspective of a change in years of schooling but rather from the perspective of changes across different school tracks. One way to alleviate this problem while sticking to the established concept years of education is to enhance this measure by post-secondary educational activities. Another option is to run separate regressions for different school types. Both strategies were realized in this study.

## 8 Conclusion

This article estimates the causal returns to education for the population of West Germany after World War II. Simply regressing wages on a measure of education produces the well-known endogeneity bias as omitted ability variables are likely to upward shift the coefficients. One potential remedy for this problem is IV methods. The interpretation of IV estimates is bounded to the LATE, which essentially depends on the instrument. Different instruments produce different coefficients according to the characteristics of the population's subsample of compliers. This interpretation delivers the key to understanding diverging IV estimates of the returns to schooling in Germany. Ichino and Winter-Ebmer (2004) use own and father's World War II involvement as instruments and deliver estimates of 11.3% and 9.3%, respectively. Pischke and Wachter (2008) find 0% returns to schooling by instrumenting education with changes in the compulsory schooling legislation. In contrast, Becker and Siebern-Thomas (2007) report returns of 10% using the urbanization of the place of birth as an instrument. My study reconciles these seemingly conflicting results. Using SOEP data, I estimate the returns to schooling for each of the corresponding subsamples that presumably represents the compliers of each instrument. These samples are (A) all available individuals, (B) graduates from the basic school track, and (C) graduates from the higher school tracks, respectively. My calculations confirm the directions of the studies' results and hence contribute to externally validate them.

I use a control function approach to regress averaged log hourly wages on a measure of education that includes post-secondary education and a set of control variables. Identification is established through the nonlinearity of the control term, which is driven by heteroskedasticity. Tentative evidence is provided that the necessary conditions for deriving causal estimates, the variable impact property (VIP) and the constant correlation condition (CCC), are likely to hold. Arguing that the presence of heteroskedasticity is not bounded to a subsample of my observations, the final coefficients sidestep the limited LATE interpretation while instead possessing ATE interpretation. For the entire sample, one additional year of education increases wages by 8.5%. The CF approach reveals that the OLS parameter is only slightly downward biased. The picture changes for the subsample of basic school

track individuals. For them, one additional year of education increases wages by 2.3% after controlling for endogeneity. The OLS estimate of 4.4% is strongly upward biased, which is in line with the notion of omitted ability variable bias. In stark contrast, for the subsample of higher school tracks graduates, the estimated return to education is 11%. The corresponding OLS estimate of 7.1% is downward biased. The robustness section shows that these results are fairly robust across different specifications and sample definitions. Splitting the samples in two across time, however, reveals that the estimations are most reliable for early born individuals that belong to the birth cohorts from 1939 until the late 1950s. Finally, assessing the impact of measurement error in the wage variable strengthens the argument for employing time-averaged observations. Potential measurement error in the education variable leads to the conclusion that the CF estimates are likely to represent a lower bound of the true parameter.

## References

- Acemoglu, Daron and Joshua D. Angrist (2000). “How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws”. In: *NBER/ Macroeconomics Annual* 15.1, pp. 9–59.
- Angrist, Joshua D. and Alan B. Krueger (1991). “Does Compulsory School Attendance Affect Schooling and Earnings?” In: *Quarterly Journal of Economics* 106.4, pp. 979–1014.
- Becker, Gary S. (1991). *A Treatise on the Family*. 2nd ed. Harvard University Press.
- (1993). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. 3rd ed. Chicago: The University of Chicago Press.
- Becker, Sascha and Frank Siebern-Thomas (2007). *Schooling Infrastructure, Educational Attainment and Earnings*. Updated version of: Returns to Education in Germany: A variable treatment intensity approach. EUI Working Paper ECO (9), 2001.
- Black, Dan A., Marc C. Berger, and Frank A. Scott (2000). “Bounding Parameter Estimates with Nonclassical Measurement Error”. In: *Journal of the American Statistical Association* 95.451.
- Bound, John, Charles Brown, and Nancy Mathiowetz (2001). “Measurement Error in Survey Data”. In: *Handbook of Econometrics*. Ed. by James J. Heckman and E. Leamer. Vol. 5. Amsterdam: Elsevier Science, pp. 3705–3843.
- Bound, John and Gary Solon (1999). “Double Trouble: On the Value of Twins-Based Estimation of the Return to Schooling”. In: *Economics of Education Review* 18, pp. 169–182.
- Card, David (1995). “Using Geographic Variation in College Proximity to Estimate the Return to Schooling”. In: *Aspects of Labour Market Behaviour*. Ed. by Loizos N. Christofides, E. Kenneth Grant, and Robert Swidinsky. Toronto and Buffalo: University of Toronto Press, pp. 201–222.
- (1999). “The Causal Effect of Education on Earnings”. In: *Handbook of Labor Economics* 3, pp. 1801–1863.

- Conneely, Karen and Roope Uusitalo (1998). *Estimating Heterogeneous Treatment Effects in the Becker Schooling Model*. Department of Economics, University of Helsinki.
- Devereux, Paul J. and Robert A. Hart (2010). “Forced to be Rich? Returns to Compulsory Schooling in Britain”. In: *The Economic Journal* 120.549, pp. 1345–1364.
- Dustmann, Christian (2004). “Parental Background, Secondary School Track Choice, and Wages”. In: *Oxford Economic Papers* 56.2, p. 209.
- Farré, Lydia, Roger Klein, and Francis Vella (2009). “Does Increasing Parents’ Schooling Raise the Schooling of the Next Generation? Evidence Based on Conditional Second Moments”. In: *IZA Discussion Paper* 3967.
- (2010). “A Parametric Control Function Approach to Estimating the Returns to Schooling in the Absence of Exclusion Restrictions. An Application to the NLSY”. In: *IZA Discussion Paper* 4935.
- Gang, Ira N. and Klaus F. Zimmermann (2000). “Is Child Like Parent? Educational Attainment and Ethnic Origin”. In: *Journal of Human Resources* 35.3, pp. 550–569.
- Griliches, Zvi (1977). “Estimating the Returns to Schooling: Some Econometric Problems”. In: *Econometrica* 45.1, pp. 1–22.
- Harmon, Colm and Ian Walker (1995). “Estimates of the Economic Return to Schooling for the United Kingdom”. In: *The American Economic Review* 85.5, pp. 1278–1286.
- Ichimura, Hidehiko (1993). “Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models”. In: *Journal of Econometrics* 58.1-2, pp. 71–120.
- Ichino, Andrea and Rudolf Winter-Ebmer (2004). “The Long-Run Educational Cost of World War II”. In: *Journal of Labor Economics* 22.1.
- Imbens, Guido W. and Joshua D. Angrist (1994). “Identification and Estimation of Local Average Treatment Effects”. In: *Econometrica*, pp. 467–475.
- Kane, Thomas J. and Cecilia E. Rouse (1995). “Labor-Market Returns to Two- and Four-Year College”. In: *The American Economic Review* 85.3, pp. 600–614.



- Kane, Thomas J., Cecilia E. Rouse, and Douglas Staiger (1999). “Estimating Returns to Schooling when Schooling is Misreported”. In: *NBER Working Paper* 7235.
- Klein, Roger and Francis Vella (2009). “Estimating the Return to Endogenous Schooling Decisions via Conditional Second Moments”. In: *Journal of Human Resources* 44.4, pp. 1047–1065.
- (2010). “Estimating a Class of Triangular Simultaneous Equations Models Without Exclusion Restrictions”. In: *Journal of Econometrics* 154.2, pp. 154–164.
- Kling, Jeffrey R. (2001). “Interpreting Instrumental Variables Estimates of the Returns to Schooling”. In: *Journal of Business and Economic Statistics* 19.3, pp. 358–364.
- Mincer, Jacob (1974). *Schooling, Experience, and Earnings*. New York: Columbia University Press.
- Oreopoulos, Philip (2006). “Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter”. In: *American Economic Review* 96.1, pp. 152–175.
- (2007). “Do Dropouts Drop Out Too Soon? Wealth, Health and Happiness from Compulsory Schooling”. In: *Journal of Public Economics* 91.11-12, pp. 2213–2229.
- Pischke, Jörn-Steffen (2007). “The Impact of Length of the School Year on Student Performance and Earnings: Evidence from the German Short School Year”. In: *The Economic Journal* 117.523, pp. 1216–1242.
- Pischke, Jörn-Steffen and Till von Wachter (2008). “Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation”. In: *Review of Economics and Statistics* 85.3, pp. 1278–1598.
- Racine, Jeffrey S. (2008). “Nonparametric Econometrics: A Primer”. In: *Foundations and Trends in Econometrics* 3.1, pp. 1–88.
- Schroeder, Elizabeth (2010). *The Impact of Microcredit Borrowing on Household Consumption in Bangladesh*. Department of Economics, Georgetown University.

Siedler, Thomas (2010). “Schooling and Citizenship in a Young Democracy: Evidence from Postwar Germany”. In: *Scandinavian Journal of Economics* 112.2, pp. 315–338.

Silverman, Bernard W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

Solon, Gary (1992). “Intergenerational Income in the United States”. In: *American Economic Review* 82.3, pp. 393–408.

## A Appendix

### A.1 The Klein and Vella (2010) Approach

Klein and Vella’s (2010) approach start with the controlled function

$$W = X\beta + \delta S + \lambda v + e. \tag{A-1}$$

Its endogeneity parameter  $\lambda$  comes from the errors equation  $u = \lambda v + e$  and can be decomposed as following:

$$\lambda = \frac{\text{cov}(u, v)}{\text{var}(v)} = \frac{\text{cov}(u, v)}{\sigma_u \sigma_v} \frac{\sigma_u}{\sigma_v} = \rho \frac{\sigma_u}{\sigma_v} \tag{A-2}$$

where  $\sigma_j$ ,  $j = u, v$ , denotes the standard deviations of the error terms  $u$  and  $v$ , and  $\rho = \text{cov}(u, v)/\sigma_u \sigma_v$  is the correlation coefficient between them. Next assume that the errors are heteroskedastic: Their distributions — given by the standard deviations  $\sigma_j$  — are a function of  $X_j$ . Let  $H_j(X_j)$  be such heteroskedasticity function. The control term impact now becomes  $\lambda(X_u, X_v) = \rho_{uv} [H_u(X_u)/H_v(X_v)]$  and is no longer constant. Plugging this conditional variant of  $\lambda$  into (A-1) yields the final estimation equation:

$$W = X\beta + \delta S + \rho \frac{H_u(X_u)}{H_v(X_v)} \hat{v} + \varepsilon. \tag{A-3}$$

To support the plausibility of the *constant correlation condition* (CCC), let us decompose the errors to a multiplicative structure. The terms  $u^*$  and  $v^*$  represent

the unscaled parts of the errors with constant variance.  $H_j(X_j)$  represents the heteroskedastic parts of the errors

$$u = H_u(X_u)u^* \quad (\text{A-4})$$

$$v = H_v(X_v)v^*. \quad (\text{A-5})$$

The correlation coefficient  $\rho$  can now be written as

$$\rho = \frac{\text{cov}(u, v)}{H_u(X_u)H_v(X_v)} = \frac{H_u(X_u)H_v(X_v)\text{cov}(u^*, v^*)}{H_u(X_u)H_v(X_v)} = \text{cov}(u^*, v^*) \quad (\text{A-6})$$

This reformulation shows that the CCC depends on a constant degree of endogeneity across  $X$  *independent* of the presence of heteroskedasticity.

## A.2 Three-Step Estimation Procedure

The estimation procedure follows a three-step sequence that is detailed in the following.

(1) OLS of education equation:  $\hat{v} = S - X\hat{\phi}_{OLS}$

(2) SLS of education equation:  $\hat{H}_v(X_v) = \hat{E}[\hat{v}^2|X_v\hat{\pi}]$  with  
 $\hat{\pi} = \arg \min_{\pi} \sum(\hat{v}^2 - \hat{E}[\hat{v}^2|X_v\pi])$

(3a) OLS of wage equation with incomplete control term:

$$\hat{u}(\Phi^*) = W - X\hat{\beta}_{OLS} - \hat{\delta}_{OLS}S - \hat{\rho}_{OLS}\frac{1}{H_v(X_v)}\hat{v}$$

(3b) SLS of wage equation:  $\hat{H}_u(X_u) = \hat{E}[\hat{u}(\Phi^*)^2|X_u\hat{\theta}]$  with

$$\hat{\theta} = \arg \min_{\theta} \sum(\hat{u}^2 - \hat{E}[\hat{u}(\Phi^*)^2|X_u\theta])$$

(3c) Complete control term to  $\frac{\hat{H}_u(X_u)}{\hat{H}_v(X_v)}\hat{v}$  and run steps (3a) and (3b) iteratively until all parameters converge.

The standard errors are obtained by bootstrapping across all steps from (1)

to (3c). This iteration procedure was first introduced by Farré, Klein, and Vella (2010). In their primary article KV (2010) propose maximum likelihood methods (ML) instead of simple iterations for steps 3a-c. While ML is less prone to local optima and corner solutions it is computationally infeasible in my application.

### **A.3 Practical Aspects of Implementing the Estimator**

The estimator was programmed and executed in R, a software environment for statistical computing. For the SLS estimation, I used the pre-built module `NPINDEX` which is available in the NP software package. Applying the software to the above outlined estimator requires various choices made by the researcher. First and foremost, the chosen method brings about a trade-off between accuracy and computation time. Note that while SLS is computationally less demanding than fully nonparametric regressions, large datasets and many regressors may render its computation burdensome nonetheless. For the entire estimation procedure, this computational demand multiplies with the number of iterations times the number of bootstrap replications. Altogether this quickly exceeds the capacity of typical workstations. To nonetheless adopt settings that grant reasonably accurate estimations, the results of this paper were produced using high-performance parallel processing, that is, utilizing up to 500 computer cores at the same time. In order to keep overall computation time practical, the number of times the `NPINDEX` routine restarts the cross-validation process of finding local extrema is set to 3. Its internal relative convergence tolerance is set to 0.001%. The nonparametric fitting employs a second order Epanechnikov kernel based on Silverman's rule-of-thumb bandwidth (Silverman 1986). Iterative convergence in step 3 of the estimation procedure is defined to be accomplished when three successive iterations produce point estimates that deviate 0.01% or less from one another. Convergence is considered to fail after 50 iterations. Monte Carlo simulations (not reported in this paper) reveal that the chosen tolerance margins produce consistent estimates that do not significantly differ from more rigorous settings.

Table A-I  
HETEROSKEDASTICITY INDEX — EDUCATION EQUATION

Dependent variable: squared residuals from OLS education equation						
	ALL		BASIC TRACK		HIGHER TRACKS	
	$\beta_{SLS}$	s.e.	$\beta_{SLS}$	s.e.	$\beta_{SLS}$	s.e.
BIRTH	1.000	(0.000)	1.000	(0.000)	1.000	(0.000)
FEMALE	-0.456	(0.233)	0.130	(0.068)	-0.404	(0.202)
RURAL	-1.432	(0.280)	-0.036	(0.067)	-0.980	(0.276)
MIGBACK	-1.044	(0.509)	0.706	(0.125)	0.425	(0.509)
AGEIMMIG	-0.335	(0.001)	0.081	(0.000)	0.204	(0.001)
SIBLINGS	-0.571	(0.009)	0.057	(0.002)	-0.244	(0.001)
Schleswig-Holstein	-1.732	(0.472)	-0.001	(0.096)	-0.822	(0.451)
Hamburg	-0.332	(0.622)	-0.471	(0.395)	-0.004	(0.866)
Lower Saxony	-0.718	(0.271)	-0.090	(0.099)	-0.532	(0.227)
Bremen	1.678	(1.033)	0.497	(0.077)	1.896	(0.648)
Hesse	0.110	(0.338)	-0.095	(0.109)	-0.047	(0.229)
North Rhine-Westphalia	-0.641	(0.320)	-0.092	(0.163)	-0.269	(0.286)
Baden-Württemberg	-0.348	(0.269)	-0.017	(0.068)	0.062	(0.243)
Bavaria	-0.953	(0.260)	-0.190	(0.104)	-0.430	(0.129)
n	21,558		9,931		11,627	
R <sup>2</sup>	0.66		0.84		0.68	

Note: SOEP, 1984–2009. Cluster robust standard errors in parentheses. The sample includes person-year observations for all persons employed in West-Germany (excluding self-employed and West Berlin) born after 1939, and aged between 30 and 65 years, for whom all relevant variables are non-missing. Estimated via semiparametric least squares (SLS). See Ichimura (1993) for the calculation of the variance-covariance matrix. The goodness-of-fit measure  $R^2 = [\sum(y_i - \bar{y})(\hat{y}_i - \bar{y})]^2 / \sum(y_i - \bar{y})^2 \sum(\hat{y}_i - \bar{y})^2$  is an adaption of the standard  $R^2$  to the non-parametric environment. It lies in the range  $[0, 1]$  with the value 1 denoting a perfect fit (Racine 2008).

Table A-II  
HETEROSKEDASTICITY TEST —  
EDUCATION EQUATION: ENTIRE SAMPLE

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity		
Ho: Constant variance		
	Chi <sup>2</sup>	p
BIRTH	2.840	0.092
FEMALE	16.818	0.000
RURAL	199.533	0.000
MIGBACK	111.379	0.000
AGEIMMIG	101.777	0.000
SIBLINGS	218.307	0.000
Schleswig-Holstein	9.998	0.002
Hamburg	1.255	0.263
Lower Saxony	3.810	0.051
Bremen	14.422	0.000
Hesse	9.678	0.002
North Rhine-Westphalia	6.833	0.009
Baden-Württemberg	10.503	0.001
Bavaria	17.564	0.000
Simultaneous	532.36	0.000

Table A-III  
HETEROSKEDASTICITY INDEX — WAGE EQUATION

Dependent variable: squared residuals from OLS wage equation						
	ALL		BASIC TRACK		HIGHER TRACKS	
	$\beta_{SLS}$	s.e.	$\beta_{SLS}$	s.e.	$\beta_{SLS}$	s.e.
YRSEDUC	1.000	(0.000)	1.000	(0.000)	1.000	(0.000)
BIRTH	0.011	(0.045)	0.005	(0.023)	0.019	(0.043)
AGE	0.002	(0.002)	0.000	(0.002)	0.007	(0.003)
AGE <sup>2</sup>	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
TENURE	-0.002	(0.001)	0.000	(0.000)	-0.008	(0.001)
FEMALE	-0.016	(0.296)	0.001	(0.252)	-0.039	(0.356)
MIGBACK	0.000	(0.312)	-0.001	(0.169)	-0.006	(1.128)
PUBLIC	-0.015	(0.411)	-0.022	(0.192)	-0.035	(0.396)
FULLTIME	-0.017	(0.884)	-0.040	(0.767)	0.002	(1.166)
EXPER	0.000	(0.001)	0.000	(0.001)	-0.003	(0.001)
UNEMPL	0.000	(0.022)	0.001	(0.005)	-0.014	(0.250)
Schleswig-Holstein	0.000	(0.583)	0.007	(0.177)	0.000	(1.148)
Hamburg	0.005	(0.936)	-0.002	(0.542)	0.011	(1.276)
Lower Saxony	0.012	(0.361)	-0.004	(0.224)	0.021	(0.323)
Bremen	-0.035	(0.793)	0.006	(0.759)	-0.039	(0.868)
Hesse	0.000	(0.415)	-0.001	(0.329)	0.029	(0.350)
North Rhine-Westphalia	-0.006	(0.445)	-0.011	(0.344)	-0.013	(0.391)
Baden-Württemberg	-0.009	(0.448)	-0.006	(0.179)	0.007	(0.361)
Bavaria	0.000	(0.393)	-0.003	(0.183)	-0.008	(0.363)
agriculture	0.020	(1.131)	-0.007	(0.649)	0.019	(0.430)
energy	0.013	(0.675)	0.031	(0.592)	-0.005	(1.276)
mining	0.031	(0.735)	0.078	(0.590)	-0.032	(1.447)
construction	-0.017	(0.737)	-0.012	(0.279)	-0.037	(0.547)
trade	0.012	(0.490)	-0.007	(0.216)	0.026	(0.349)
transport	0.002	(0.467)	0.002	(0.196)	0.016	(0.637)
banking & insurance	0.019	(0.431)	0.031	(0.306)	0.011	(0.339)
services	0.002	(0.393)	0.010	(0.289)	-0.019	(0.383)
other industries	0.017	(1.281)	-0.003	(1.283)	0.074	(2.221)
20 – 200 employees	-0.004	(0.430)	-0.009	(0.199)	-0.012	(0.387)
200 – 2000 employees	-0.016	(0.419)	-0.022	(0.202)	-0.028	(0.523)
> 2000 employees	-0.009	(0.468)	-0.020	(0.181)	-0.007	(0.384)
n	21, 558		9, 931		11, 627	
R <sup>2</sup>	0.26		0.14		0.29	

Note: See note of Table A-I.

Table A-IV  
HETEROSKEDASTICITY TEST — WAGE EQUATION: ENTIRE SAMPLE

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity		
Ho: Constant variance		
	Chi <sup>2</sup>	p
YRSEDUC	196.55	0.000
BIRTH	1.52	0.218
AGE	58.42	0.000
AGE <sup>2</sup>	59.39	0.000
TENURE	231.79	0.000
FEMALE	57.34	0.000
MIGBACK	13.57	0.000
PUBLIC	131.80	0.000
FULLTIME	253.62	0.000
EXPER	88.89	0.000
UNEMPL	12.79	0.000
Schleswig-Holstein	1.11	0.293
Hamburg	6.86	0.009
Lower Saxony	51.44	0.000
Bremen	5.42	0.020
Hesse	0.22	0.636
North Rhine-Westphalia	16.86	0.973
Baden-Württemberg	23.08	0.000
Bavaria	0.00	0.973
agriculture	8.10	0.004
energy	1.11	0.292
mining	2.50	0.114
construction	81.78	0.000
trade	120.41	0.000
transport	8.66	0.003
banking & insurance	25.76	0.000
services	0.09	0.000
other industries	2.43	0.119
20 – 200 employees	18.98	0.000
200 – 2000 employees	67.28	0.000
> 2000 employees	31.17	0.000
Simultaneous	1,387.76	0.000



Figure A-I  
 EDUCATION EQUATION — GRAPHICAL ANALYSIS OF  
 HETEROSKEDASTICITY:ENTIRE SAMPLE

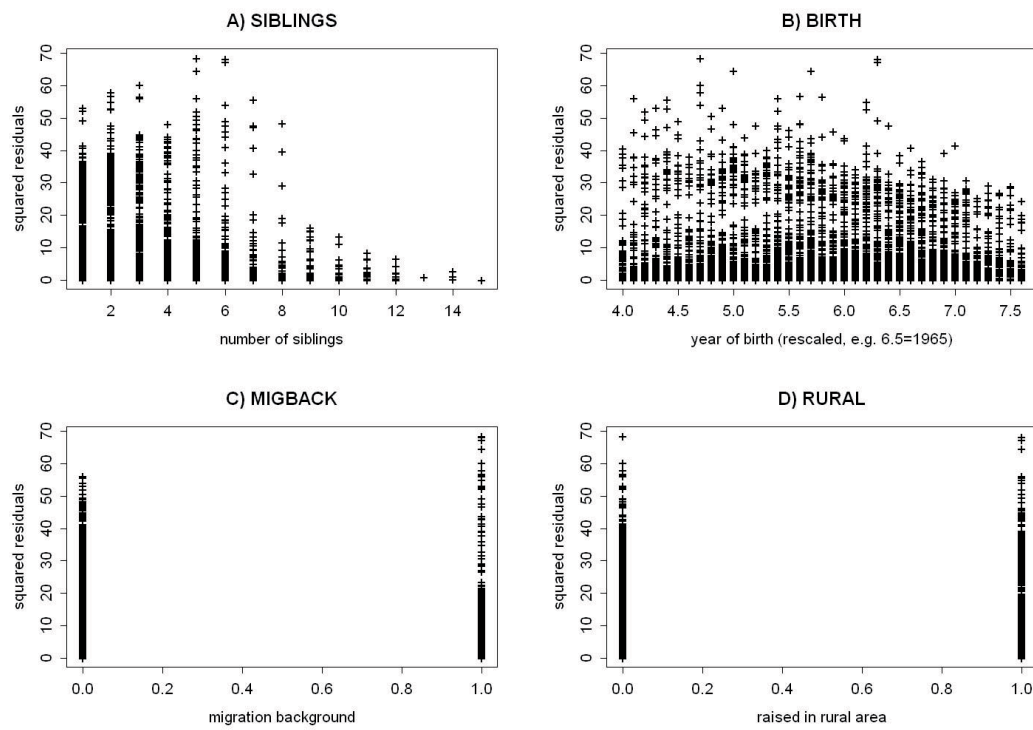


Figure A-II  
WAGE EQUATION — GRAPHICAL ANALYSIS OF  
HETEROSKEDASTICITY:ENTIRE SAMPLE

