The Potential Outcome of Schooling: Individual Heterogeneity, Program Risk and Residual Wage Inequality

Winfried Pohlmeier *
University of Konstanz
CoFE, ZEW

Anton L. Flossmann
University of Konstanz

Very preliminary and incomplete version
Paper to be presented at the conference "Heterogenous Labor and Education" DFG Research Group 454, Mannheim 12-13 September 2005
this version: September 8, 2005

Abstract

This paper analyzes the link between individual heterogeneity, program risk, and outcome inequality within the potential outcome framework. Using standard ignorability conditions we derive identifying conditions for the variance of the treatment effect, i.e. the program risk, and relate this parameter to the variance treatment effect. Moreover, for a rather general set-up allowing for binary as well as continuous treatments, we derive upper and lower bounds for the program risk.

Two applications demonstrate the approach at work. Using the Dehejia/Wahba (1999) data we check whether the estimates obtained by our approach can replicate the treatment parameters obtained from experimental data. Secondly, we estimate the returns from upper secondary graduation in Germany and the implied residual wage inequality.

JEL classification: $C21, J24, J31$

Keywords: correlated random coefficient model, ignorability, potential outcome approach, residual wage inequality, program risk

---

*Corresponding author. Department of Economics, Box D124, University of Konstanz, 78457 Konstanz, Germany. Phone ++49-7531-88-2660, Fax -4450, email: winfried.pohlmeier@uni-konstanz.de. Financial support by the DFG through research group 'Heterogenous Labor" at the University of Konstanz and the ZEW, Mannheim, is gratefully acknowledged.

1 Introduction

Since the seminal contributions by Becker (1964) and Mincer (1974) the analysis of the returns to schooling has been in the focus of labor economic research. While early empirical studies focused on the correlation between unobserved heterogeneity (ability) to remove the ability bias, the more recent contributions motivated by the econometrics of evaluation emphasize the endogeneity of schooling in the light of heterogeneous agents.[1] Provided that individuals differ in their (at least partly unobservable) marginal costs and benefits of schooling, the educational choice leads to heterogeneous returns to schooling. Thus selfselection into schooling levels effects not only mean return rates but the also the observed residual wage distribution. Therefore educational policies which have an impact on the individuals' marginal cost and benefit not only effect the returns to schooling, but also the residual wage inequality. The potential selectivity bias in variances is usually neglected in the literature on residual wage inequality (e.g. Juhn, Murphy, and Pierce (1993) and Katz and Autor (1999)).

In addition, individual heterogeneity also has serious implications for the quality of educational policies. Evaluating the causal effects of schooling by means of the potential outcome approach, taking into account heterogeneity of returns and the endogeneity of schooling decisions, usually focuses on mean causal effects. This ignores that the effectiveness of educational policies also depends on the riskiness of the program, i.e. the variance of the causal effect. It is needless to stress that knowledge of the program risk is another valuable dimension which risk averse policy makers are concerned about in assessing the quality of a program. The variance of causal effects, however, is not identified given the standard assumptions of the potential outcome approach.

The purpose of the paper is twofold. First, we try to shed more light on causal residual income inequality due to schooling using the potential outcome approach. Unlike Chen (2004), whose analysis is based on the Heckmans's control function approach, we use the ignorability (unconfoundedness) assumptions to identify the causal effects and base our analysis on the random coefficient specification of the earnings function. This leaves enough flexibility to analyze the nonparametric case

[1]See for example Blundell, Dearden, and Sianesi (2005) for an application as well as Card (2001) for a simple theoretical motivation.

of binary treatment (participation in a schooling program or not) and the case continuous or ordered treatments implicitly assumed in traditional earnings functions.[2]

Secondly, since causal program risk cannot be identified by the usual ignorability assumptions we derive identifying conditions for the variance of the treatment effect. Moreover, we derive bounds for the program risk, which can be nonparametric without additional identifying assumptions.

Our paper is organized as follows. In Section 2 we develop the potential outcome approach for the random coefficient model based on appropriate ignorability (unconfoundedness) assumptions. Following Wooldridge (2004), we identify the average treatment effect via conditional mean independence assumptions and show that the ATE for the continuous treatment variable or a binary treatment can be estimated by means of auxiliary regressions. For the random coefficient model we derive the bounds for the causal variance effects using additional conditional independence assumptions only and relate those bounds to the identifiable causal wage inequality. In Section 3 we describe how to estimate the variance effects under unconfoundedness using standard matching approaches. The simple way of checking the reliability of nonexperimental evaluation estimators is to confront their estimates with the ones obtained from experimental data. We do this by estimating the variance effects using the data from LaLonde (1986). Finally, in Section 4 we show the estimator at work by evaluating the causal effect of schooling for graduates from the German Gymnasium using cross-sectional data from the German Socioeconomic Panel. Section 5 concludes and gives an outlook on future research.

[2]See also Abadie (2002), Angrist (2004) and Chernozhukov and Hansen (2001) who analyze quantile treatment effects using nonparametric methods.

2 Identifying Treatment Parameters under Ignorability

Our general starting point is the standard correlated random coefficient model of the form

$$Y = \alpha + \beta S, \tag{2.1}$$

where Y is the outcome variable (e.g. log income) and α and β are correlated random coefficients. The scalar treatment variable S (schooling) can be continuous, a count, or a binary treatment variable. The term "correlated" refers to the property that the two coefficients are random variables correlated with attributes affecting the outcome variable through α and β. It is needless to stress that the treatment variable is endogenous and is also correlated with unobservable factors and observable attributes. Note that (2.1) is general enough to capture a variety of specifications. The classic Becker-Mincer earnings function arises with $\alpha = X'\alpha_0 + \varepsilon$ and $\beta = \beta_0$, where α_0 and β_0 are fixed coefficients, X a vector of attributes (experience, etc.) and ε an error term capturing unobserved abilities possibly correlated with schooling. Moreover, if β is correlated with observable and unobservable factors, the specification corresponds to the one proposed by Garen (1984) and Heckman and Vytlacil (1998).

Mean Effects

Wooldridge (2004) proposes an estimation approach for the average partial effect $\Delta_{APE} := \mathrm{E}\left[\frac{\partial Y}{\partial S}\right] = \mathrm{E}[\beta]$ based on the following conditional mean independence assumptions (ignorability conditions) as identification strategy.[3]

Assumption 2.1 (Ignorability I)
Let Y be the outcome variable and S the treatment. For a set of covariates X the following conditions hold:

i. The relationship between outcome and treatment is given by the random coefficient model (2.1).

ii. $\mathrm{E}[Y|\,\alpha, \beta, S, X] = \mathrm{E}[Y|\,\alpha, \beta, S]$

iii. Conditional on X, α and β are redundant in the first two conditional moments of S: a) $\mathrm{E}[S|\,\alpha, \beta, X] = \mathrm{E}[S|\,X]$ *and b)* $\mathrm{V}[S|\,\alpha, \beta, X] = \mathrm{V}[S|\,X] > 0$

[3]See also the textbook by Wooldridge (2002, pp. 639) for a brief description of this approach.

Identification condition *ii* obviously holds since the control variable X enters the equation through α, β and S only. Assumption *iii.a)* guarantees that, conditional on the controls, expected treatment is mean independent of α and β. Thus no new information is gained in projecting treatment if there are sufficient controls. Assumption *iii.b)* is closely related to *iii.a)* and extends the ignorability assumption to the second moments of S. As shown in the Appendix, these assumptions build crucial identification conditions (ignorability conditions) needed to identify the average partial effect Δ_{APE}.[4]

Proposition 2.1 (APE in the RC-Model under Ignorability)
Given the ignorability assumptions ii) and iii) the average partial effect Δ_{APE} of the random coefficient model 2.1 is given by:

$$\mathrm{E}[\beta] = \mathop{\mathrm{E}}_{X}[\mathrm{E}[\beta|X]] = \mathop{\mathrm{E}}_{X}\left[\frac{\mathrm{Cov}[S, Y|X]}{\mathrm{V}[S|X]}\right]. \tag{2.2}$$

The similarity of equation (2.2) to the linear predictor formula is by no means incidental. In fact, Wooldridge (2004) derives Δ_{APE} as the expectation of the linear predictor of Y on S conditional on X. For the case of a binary treatment $S \in \{0, 1\}$. The random coefficient model (2.1) with ignorability assumptions *ii* and *iii a)* is simply an alternative representation of the potential outcome model under unconfoundedness. In this case $\alpha = Y_0$ and $\beta = Y_1 - Y_0$, where Y_1 and Y_0 are the potential outcomes for the treatment and the nontreatment case. Thus β is the conditional average treatment effect:

$$\Delta_{ATE}(X) := \mathrm{E}[Y_1 - Y_0|\, X] = \mathrm{E}[\beta|\, X].$$

Without loss of generality, the ignorability conditions are reversed compared to the literature on estimation of binary treatments under unconfoundedness. There, the conditional mean independence assumption is defined in terms of the mean of the outcome variable conditional on the treatment indicator and the controls, while in Assumption 2.1 unconfoundedness is defined in terms of the conditional mean of the treatment variable.

[4] Since the approach is general enough to deal with discrete and continuous treatments we prefer to use the term average partial effect rather than average treatment effect used in the literature on binary treatment effects.

Moreover, note that the binary treatment case (2.1) is fully nonparametric and does not impose any functional form restrictions. The ignorability condition *iii.b)* becomes redundant, since the conditional variance of the treatment variable is merely determined by the mean function. Let $\mu_S(X) \equiv \mathrm{E}[S|\,X]$ and $\omega_S^2(X) \equiv \mathrm{V}[S|\,X]$ then (2.2) can be rewritten as

$$\mathrm{E}[\beta|X] = \frac{\mathrm{E}[(S-\mu_S(X))Y|\,X]}{\omega_S^2(X)}. \tag{2.3}$$

Since we are assuming an iid random sample a consistent estimator for $\mathrm{E}[\beta]$ is given by

$$\hat{\mathrm{E}}[\beta] = \frac{1}{n}\sum_{i=1}^{n} \frac{(S_i - \hat{\mu}_s(X_i))Y_i}{\hat{\omega}_s^2(X_i)}, \tag{2.4}$$

where $\hat{\mu}_S(X_i)$ and $\hat{\omega}_S^2(X_i)$ are estimators of the conditional mean and the variance function. Note, the $S_i - \hat{\mu}_S(X_i)$ is simply the residual of a regression of S on X. For a binary treatment, the conditional mean function of S is the propensity score, $\mu_S(X) = p(X)$ with the conditional variance function $\omega_S^2(X) = p(X)(1-p(X))$.

$$\hat{\mathrm{E}}[\beta] = \frac{1}{n}\sum_{i=1}^{n} \frac{(S_i - \hat{p}(X_i))Y_i}{\hat{p}(X_i)(1-\hat{p}(X_i))} \tag{2.5}$$

Note that this is the feasible version of the weighting estimator (Imbens (2003)). Lemma 2.1 provides interesting implications for the generality of the random coefficient model (see the Appendix for the proof):

Lemma 2.1 (Conditional Uncorrelatedness of S and β)
Under the ignorability conditions given in Assumption 2.1, S and β conditional on X are uncorrelated:

$$\mathrm{Cov}[\beta, S|X] = 0.$$

Moreover, this implies that $\mathrm{E}[\alpha]$ is identified.

Conditional on the attributes of X the partial effect and the treatment variable are uncorrelated, i.e. given sufficient controls the model rules out nonlinearities between earnings and schooling. However, the linear form of the random coefficient model does not exclude decreasing or increasing returns to schooling. For the binary case Lemma 2.1 reflects a well-known property: since $\mathrm{Cov}[Y_1 - Y_0, S|X] = 0$ simply states that, conditional on X, the average effects for the treated, the non-treated, and the average treatment effect are identical, i.e. knowledge of the attributes corrects for selectivity.

The identifiability of α turns out to be important if $S = 0$ is an important benchmark value, e.g. if the treatment variable is binary on a non-treatment situation or baseline treatment level (minimum schooling) can be defined. In this case α is the potential outcome for the minimum treatment situation.

Variance Effects and Variance Bounds

Mean program effects may not be the only measures policy makers are interested in if they are risk averse. They also need information on the variability of a program effect, shortly its risk. However, the usual ignorability assumptions as given above are not sufficient to identify the variance of the partial effect ($\Delta_{VPE} := \mathrm{V}\left[\frac{\partial Y}{\partial S}\right] = \mathrm{V}[\beta]$). In order to derive identification conditions for higher moments we replace the conditional mean independence assumptions *iii* by the following somewhat more restrictive conditional independence assumption:

Assumption 2.2 (Ignorability II)
Let Y be the outcome variable and S the treatment. For a set of covariates X the following independence property holds:

iii. $S \perp \alpha, \beta \mid X$.

For the binary treatment case it is easy to show that this additional ignorability condition is sufficient to identify the variances of the two potential outcomes $\mathrm{V}[Y_0]$ and $\mathrm{V}[Y_1]$. Corresponding to the quantile treatment effects literature, we may call $\mathrm{V}[Y_1] - \mathrm{V}[Y_0]$ the variance treatment effect, while $\mathrm{V}[Y_1 - Y_0]$ is the variance of the treatment effect. In the case of the earnings function the two variances determine the wage inequality of two groups of individuals with different levels of education if the individuals had been selected randomly into the two groups. The difference between the two variances can be thought of as being a pure measure of residual wage inequality that is independent of the selfselection process of the residual wage distributions, respectively and thus residual wage inequality. However, Assumption 2.2 is not sufficient to identify program risk. This requires additional information on $\mathrm{Cov}[\alpha, \beta|X]$.

Proposition 2.2 (Identification of the Δ_{VPE}) *Given the ignorability assumption 2.2 the variance of the treatment effect (Δ_{VPE}) is identified, if* $\mathrm{Cov}\left[\alpha, \beta|X\right]$ *is identified.*

The proof is given in the Appendix 5. The covariance term indicates whether there are increasing or decreasing returns to treatment. Assume the treatment variable is bounded from below at zero ($\min S = 0$, e.g. required years of schooling) then α reflects outcome in the case of minimal treatment. Thus $\mathrm{Cov}\left[\alpha, \beta|X\right] = \mathrm{Cov}\left[Y_0, \beta|X\right]$ contains the information whether individuals with higher outcomes in case of non-treatment are expected to reveal higher returns than those with lower outcomes if not treated. Thus $\mathrm{Cov}\left[\alpha, \beta|X\right] < 0 (> 0)$ simply reflects decreasing (increasing) returns to treatment. Since we can give this covariance an economic interpretation we may be able to use external information (e.g. from experimental studies) to infer on the sign of $\mathrm{Cov}\left[\alpha, \beta|X\right]$ in the study of interest.

For the case of a binary treatment variable information on the covariance between α and β is equivalent to information on the correlation between the potential outcomes Y_0 and Y_1. The relationship between program risk and residual wage inequality becomes evident by reformulating the definition of the variance of the program effect in terms of the variances of the two potential outcomes:

$$\mathrm{V}\left[Y_1 - Y_0\right] = \mathrm{V}\left[Y_1\right] - \mathrm{V}\left[Y_0\right] - 2\,\mathrm{Cov}\left[\alpha, \beta\right]. \tag{2.6}$$

Assuming nonincreasing returns to treatment the residual wage inequality serves as a lower bound for program risk:

$$\mathrm{V}\left[Y_1 - Y_0\right] \geq \mathrm{V}\left[Y_1\right] - \mathrm{V}\left[Y_0\right]. \tag{2.7}$$

Obviously this lower bound is only informative if $\mathrm{V}\left[Y_1\right] - \mathrm{V}\left[Y_0\right] > 0$. Finally, we can unambiguously conclude that if $\mathrm{V}\left[Y_1\right] - \mathrm{V}\left[Y_0\right] \leq 0$ returns to treatment are nonincreasing. For the more general random coefficient model Proposition 2.3 gives a lower bound for the program risk.

Proposition 2.3 (Lower Bound of the VPE)
Given the additional independence assumptions iv.) a lower bound for the variance

of the treatment effect (Δ_{VPE}) is given by:

$$\mathrm{V}[\beta] \geq \mathop{\mathrm{E}}_{X}\left[\frac{\mathrm{V}[Y|\,X] - \mathrm{V}[\alpha|\,X] + \mathrm{E}[S|\,X]^2}{\mathrm{E}[S^2|\,X]}\right] - \mathrm{E}[\beta]^2\,, \tag{2.8}$$

if $\mathrm{Cov}[\alpha, \beta|X] \leq 0$.

The proof of Proposition 2.3 is given in Appendix 5. Like the bound given in (2.7) the proof only exploits purely statistical properties of the variance decomposition. In a similar fashion we can also derive an upper bound for the variance of the partial effect.

Proposition 2.4 (Upper Bound of the VPE)
Given the ignorability assumptions ii and iii an upper bound for the variance of the treatment effect (Δ_{VPE}) is given by:

$$\mathrm{V}[\beta] \leq \mathop{\mathrm{E}}_{X}\left[\frac{\mathrm{V}[Y|X]}{\mathrm{V}[S|X]}\right] - \mathrm{E}[\beta]^2\,. \tag{2.9}$$

The proof of Proposition 2.4 is given in Appendix 5.

3 Performance in the Light of Experimental Data

Estimation Issues

In our applications below we will concentrate on treatment effects of the treated only.[5] Mean and variance treatment effects are estimated by conventional propensity score matching methods. Let us define $\sigma_1^2 := \mathrm{V}\,[Y_1 |\, S = 1]$ and $\sigma_0^2 := \mathrm{V}\,[Y_0 |\, S = 1]$ as the variances for the two potential outcomes for the treated. The estimate of the counterfactual variance is based on the following formula:

$$\hat{V}[Y_0|S = 1] = \hat{E}[\hat{V}[Y_0|\hat{P}, S = 1]] + \hat{V}[\hat{E}[Y_0|\hat{P}, S = 1]], \tag{3.1}$$

where $\hat{P}$ denotes the estimated propensity score $P = P(X) = \Pr\,[S = 1\,|X\,]$. The estimation procedure consists of four steps:

1. Estimate $P = \Pr\,[S = 1\,|X\,]$ by a probit or logit model.
2. Estimate $E[Y_0|\hat{P}, S = 1]$ and $E[Y_0^2|\hat{P}, S = 1]$ nonparametrically (i.e. Nearest Neighbor Matching, Kernel Matching, Local Linear Matching).
3. Compute $\hat{V}[Y_0|\hat{P}, S = 1] = \hat{E}[Y_0^2|\hat{P}, S = 1] - \hat{E}[Y_0|\hat{P}, S = 1]^2$.
4. Obtain $\hat{V}[Y_0|S = 1]$ by simply averaging and taking variances over the corresponding conditional moments.

$\hat{V}[Y_1|S = 1]$ can be estimated by the sample variance of the treated observations. Estimation is performed with both Nearest Neighbor Matching (NN) and Local Linear Matching (LLM) by using a quartic kernel. In the empirical applications we consider two different global bandwidths in order to investigate the sensitivity of the results with respect to the smoothing parameter: $h = 0.5$ and $h = 1.0$. Note, the conventional data driven selection algorithms like cross validation do not yield the optimal bandwidth in terms of minimizing the mean squared error (MSE) or the integrated mean squared error (IMSE). This is because additional smoothing takes place by averaging the estimated means over different simulated counterfactual values.[6] For the sake of an easier comparison we report the variance treatment effects for the treated in terms of a percentage difference $\tau := (\sigma_1 - \sigma_0)/\sigma_0$.

[5] In later versions of the paper we plan to extend our analysis to the case of the overall population.

[6] Froelich (2004) for example derives an MSE approximation for matching estimators of the TT in the bivariate treatment case and investigates its performance in finite samples by a plug in bandwidth choice. The reliability of this approximation turns out to be not very high and conventional cross validation bandwidth selection results to perform relatively well.

An Application to the LaLonde Data

In the following we use the job training program data, which were first analyzed by LaLonde (1986) and then formed the basis of several subsequent studies in econometric evaluation research including Heckman and Hotz (1989), Dehejia and Wahba (1999), Smith and Todd (2001, 2005) or Abadie and Imbens (2002). The reason for the widespread use of this data set is the availability of an experimental data set from the "National Supported Work Program" (NSW). The experimental data were obtained by a random assignment of treatment to eligible participants. It consists of information on earnings, treatment status, background characteristics like ethnicity or age, and also earnings before treatment. Due to the randomness of the treatment assignment, estimates based on the NSW data set can be regarded as a benchmark for nonexperimental program evaluation.

Dehejia and Wahba (1999) use the NSW data to evaluate the performance of propensity score matching methods. They conclude that the experimental results can be replicated very well by nonparametric estimates based on observational data. One of the data sets Dehejia and Wahba use is a subset of the NSW data of 185 treated units and 2490 control observations of the "Panel Study of Income Dynamics" (PSID1). Like LaLonde (1986), they also extract subsets from the PSID1 data set that resemble the treatment group in terms of single preintervention characteristics. These data sets are defined as PSID2 (all men from PSID1 who were not working when surveyed in the spring of 1976) and PSID3 (all men from PSID2 who were not working in 1975).

Using the same data we estimate the variance treatment effect of the treated Δ_{VTT} by the proposed method and compare the results to the experimental benchmark. Following Dehejia and Wahba (1999), the propensity scores are estimated by a logit model. The specification of the propensity score equation differs for each sample because it is chosen such that it balances the distribution of the covariates over both treatment groups. The estimation results are given in Table 1. Regarding the estimates of the average treatment effect of the treated Δ_{TT} we are able to replicate the results by Dehejia and Wahba (1999) exactly for the method of NN-Matching. For LLM the results for the TT are very sensitive to the chosen bandwidth. The nonexperimental estimates for the variance differential vary strongly with respect to the estimation method and the sample. For the PSID1 sample, LLM with a bandwidth of 1.0 performs best, while for the PSID2 sample NN-Matching and LLM with bandwidth 1.0 yield estimates that come close to the experimental benchmark. Finally, NN-Matching performs very well when applied to the PSID3 sample.

Table 1: Estimation Results for the Dehejia/Wahba Data

Sample	**Method**	Δ_{TT}	τ
PSID1	NN-Matching	1691 (1217)	0.1845 (0.3374)
	LLM (h=0.5)	1671 (941)	0.0933 (0.2310)
	LLM (h=1.0)	1955 (1078)	0.3404 (0.5397)
PSID2	NN-Matching	1455 (1377)	0.5451 (0.3676)
	LLM (h=0.5)	1467 (1258)	0.2543 (0.2627)
	LLM (h=1.0)	993 (1261)	0.4288 (0.5355)
PSID3	NN-Matching	1120 (1491)	0.4089 (0.3511)
	LLM (h=0.5)	1055 (1400)	0.2139 (0.3354)
	LLM (h=1.0)	710 (1366)	0.2060 (0.5622)

- Standard errors (in brackets) are bootstrapped (1000 replications)
- Experimental estimates based on NSW-data: $\tau = 1.794$, $\Delta_{VTT} = 0.4347$

The standard errors (in parenthesis) of both the estimated variance treatment effect and the mean treatment effect of the treated are computed by the bootstrap method. The bootstrapped standard errors of the mean effects considerably exceed the ones reported by Dehejia and Wahba (1999), who use the empirical standard deviation. This difference can be explained by the fact that the bootstrap standard errors also account for the estimation uncertainty generated by the propensity score estimates. Note that the estimates of the variance differentials are positive for all three samples so that the variance differential can serve as an estimate of the lower bound of the variance treatment effect.

Table 2 reports on the bound for the variance treatment effect. The estimates for the upper and the lower bound are far apart. Nevertheless the lower bound estimates are quite informative. Keeping in mind that the true program risk is even higher than our lower bound estimates, we can conclude that the lower bound compared to the mean treatment effect indicates that NSW program was not very efficient for the treated.

Table 2: Estimated Bounds for the Program Risk

Sample	Method	Lower Bound	Upper Bound
PSID1	NN-Matching	4217	50640
	LLM (h=0.5)	3180	50641
	LLM (h=1.0)	5239	50631
PSID2	NN-Matching	5981	31963
	LLM (h=0.5)	5134	31962
	LLM (h=1.0)	5140	31980
PSID3	NN-Matching	5542	38758
	LLM (h=0.5)	4486	38802
	LLM (h=1.0)	4370	38809

Estimated bounds for $\sqrt{V[Y_1 - Y_0|S=1]}$

4 An Application to Educational Choice

For our application to educational choice we use data from the 2001 wave of the German Socio Economic Panel (GSOEP). In our sample we include German workers who are full-time employed and live in West-Germany. Individuals who were at the moment of the survey self-employed, part-time employed, or in vocational training are excluded from the analysis. After eliminating all units with missing values in any of the variables considered we obtain a sample size of 1054 individuals.

Table 3: Variable Definition

Label	**Description**
LNWAGE	Log gross hourly income
STIME	Years of schooling (years)
AGE	Age (years)
AGESQ	Age squared
SIBLINGS	Number of siblings
FEDUCATION	Educational degree of father (years)
MEDUCATION	Educational degree of mother (years)
URBAN	Dummy, if individual grew up in an urban area
FPROF	Dummy for occupational position of father (1= Public servant/self employed/white-collar, 0 else)
MPROF	Dummy for occupational position of mother (1= Public servant/self employed/white-collar, 0 else)
LIVING	Dummy indicating, if the individual grew up with both parents
PINTEREST	Dummy for parental interest in educational achievement of the individual (1=strongly interested, 0 else)
MUSIC	Dummy indicating, if the individual was active in music during youth (1=active, 0 else)
SPORT	Dummy indicating, if the individual was actively doing sport during youth (1=active, 0 else)
FARGUE	Dummy for argue or fight with father, when the individual was 15
MARGUE	Dummy for argue or fight with father, when the individual was 15
Dummies for region of last school attendance (base category: North Rhine-Westphalia):	
SCHLESACHS	Schleswig-Holstein or Lower Saxony
BWBAY	Baden-Wuerttemberg or Bavaria
WBERLIN	West-Berlin
BREMHAM	Bremen or Hamburg
RPHESAAR	Rhinland-Palatinate/ Hesse /Saarland

Table 3 gives an overview of the variables and its definitions. Apart from the usual covariates on family background in human capital-earnings equations, variables that

Table 4: Summary Statistics

Variable	Mean	Std.error
LNWAGE	3.3968	.0434
STIME	12.7348	2.6967
AGE	42.0313	10.1752
SIBLINGS	1.7789	1.635
FEDUCATION	10.9967	1.471
MEDUCATION	10.5588	1.0211
URBAN	.6531	
FPROF	.4279	
MPROF	.2429	
LIVING	.9032	
PINTEREST	.6281	
MUSIC	.2448	
SPORT	.6803	
FARGUE	.2429	
MARGUE	.1034	
SCHLESACHS	.1499	
BWBAY	.3435	
WBERLIN	.0028	
BREMHAM	.0332	
RPHESAAR	.1983	
N=1054		

indicate activeness of the individual in music and sport are added to proxy individual motivation. Table 4 reports the descriptive statistics of our sample. Our treatment variable consists of two categories, reflecting the special institutional setting of the German schooling system and is formed by the highest school degree obtained: Secondary/ intermediate school (Low) and technical school/upper secondary school (High). We allow for further educational degrees like apprenticeship, foreman, university, or higher technical college. Since a technical school degree and an upper secondary school degree allow one to obtain a higher technical college degree or a university degree, we want to investigate the effect of this type of higher education on earnings. Table 5 contains some information about about the structure of qualifications. Most of the individuals have an intermediate degree while only 23 % finished the upper secondary school.

As in the application in the previous, section the propensity scores are estimated by a logit model. Regarding the relevance of a common support Heckman, Ichimura, and Todd (1997) and Heckman, Ichimura, Smith, and Todd (1998) show by comparison

Table 5: Structure of Qualifications

Highest school degree	**Frequency in %**
Secondary School	37
Intermediate or technical school	40
Upper secondary school	23
N=1054	

of experimental and nondepartmental estimation results, that an insufficient support constitutes one of the primary components of selection bias. To construct a common region of support we use a method applied in a similar way to Heckman, Ichimura, and Todd (1997) and Heckman, Ichimura, Smith, and Todd (1998) or Smith and Todd (2005). In a first step the estimated common support is obtained by:

$$\hat{\mathcal{S}}_p = \{P(x) : \hat{f}(P(x)|S=1) > 0 \ \text{ and } \ \hat{f}(P(x)|S=0) > 0\}.$$

The densities are estimated nonparametrically by kernel-densities. The optimal bandwidth is chosen by least squares cross validation. We use a quartic kernel to allow for the possibilities of estimated densities with zero values. In a second step for both $S = 1$ and $S = 0$ the observation with the lowest two percent of the estimated densities are trimmed to obtain a common region of support with densities strictly greater than zero.

In order to estimate Δ_{TT} and the variance treatment effect τ, we apply again NN-Matching and LLM. The estimated parameters are annualized by dividing the estimates by the difference between the averages in years of schooling for both groups. The results are given in Table 6.

Table 6: Estimated Average Treatment and Variance Treatment Effect

METHOD	Δ_{TT}	τ
NN-Matching	0.0418 (0.0172)	-0.0041 (0.0352)
LLM (h=0.5)	0.0417 (0.0125)	-0.0019 (0.0259)
LLM (h=1.0)	0.0461 (0.0089)	-0.0066 (0.0192)

Standard errors (in parenthesis) are bootstrapped with 500 replications.

The annualized causal return rate of school leavers with upper secondary education is between 4 and 5 percent depending on the matching method chosen. School leavers with an upper secondary degree can expect 4.2 - 4.8 per cent higher income for each school year invested compared to the counterfactual case if they had not decided to graduate from upper secondary school. In contrast to the previous application, the estimation results are not very sensitive with respect to the matching methods or bandwidth choice chosen. The bootstrapped standard errors also indicate that estimated average treatment effect of the treated is different from zero.

The estimates for the variance differential are negative but not significantly different from zero. At least for the treated we find no empirical support for differences in the residual wage inequality. Moreover, the nonpositive variance treatment effect implies that the lower bound of the variance of the treatment effect is not sharp. Thus we may conclude that the returns to schooling are nondecreasing. Without stressing this argument too much, our findings support the view proposed by Cunha, Heckman, Lochner, and Masterov (2005) that schooling enhances wage inequality in the sense that those higher unobserved skills profit more from schooling than others.

Table 7: Estimated Bounds for the Program Risk

METHOD	Upper Bound
NN-Matching	0.3992
LLM (h=0.5)	0.3992
LLM (h=1.0)	0.3992

Estimated bounds for $\sqrt{\mathrm{V}\,[V[Y_1 - Y_0|\,S = 1]}$.

The estimated upper bound of the variance of the treatment effect reported in Table 7 is computed by dividing the difference between the averages in years of schooling for both groups in order to obtain a standard deviation bound for the annualized TT. Interestingly, the estimates do not vary across estimation methods, but the upper bound is larger then the mean effect by a factor of 10, so that it contains little information for the data used here.

5 Conclusions

Based on the potential outcome approach this paper analyzes the link between individual heterogeneity, program risk, and outcome inequality. Using standard ignorability conditions we derive identifying conditions for the variance of the treatment effect, i.e. the program risk, and relate this parameter to the variance treatment effect. Moreover, for a rather general set-up allowing for binary as well as continuous treatments, we derive upper and lower bounds for the program risk.

Applying our approach to the LaLonde data we show that the lower bound for variance of the treatment effect is rather high indicating a considerable inefficiency of the NSW program. In the second application we estimate the causal effects of graduation from higher secondary school. In this application the difference of the residual income variance is negative and thus not informative. However, we find evidence for increasing returns to schooling: graduates from upper secondary school with high incomes in case of nongraduation can expect higher returns than their classmates with lower incomes.

With the potential outcome approach adopted here we are able to estimate residual wage inequality due to schooling taking into account the endogeneity of the decision process. Therefore our approach maybe used to scrutinize the empirical findings on the change of the residual wage distribution in the light of selfselection.

However, the nonparametric set-up chosen here is less informative about the sources of income variation and program risk. Ex-post observed income variation may be due to individual heterogeneity or ex-post shocks (uncertainty). Individual heterogeneity as a source of residual wage inequality emphasizes uncertainty of the econometrician about the true data generating process at the individual level. But, ex-post shocks also lead to randomness in wages and the returns to schooling. While the latter source of variation is more a question of the general macroeconomic conditions knowledge about unobserved individual, heterogeneity may help to design more efficient programs. Future work should be concerned with disentangling the two effects. This would require a more structural set-up that allows us to identify the two sources of ex-post observable income variation.

References

Abadie, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292.

Abadie, A., and G. Imbens (2002): "Simple and Bias-Corrected Matching Estimators for Average Treatment Effects," NBER Working Paper No. T0283.

Angrist, J. (2004): "Treatment Effect Heterogeneity in Theory and Practice," *The Economic Journal*, 114, C52 – C83.

Becker, G. S. (1964): *Human Capital*. Columbia University Press, New York.

Blundell, R., L. Dearden, and B. Sianesi (2005): "Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from the NCDS," *Journal of the Royal Statistical Society, Series A*, 168, 473 – 512, IFS Working Papers W03/20, Institute for Fiscal Studies.

Card, D. (2001): "Estimating the Returns to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69, 1127–1160.

Chen, S. (2004): "Estimating the Variance of Wages in the Presence of Selection and Unobserved Heterogeneity," Department of Economics, SUNY at Albany.

Chernozhukov, V., and C. Hansen (2001): "An IV Model of Quantile Treatment Effects," MIT Working Paper 02-06.

Cunha, F., J. Heckman, L. Lochner, and D. V. Masterov (2005): "Interpreting the Evidence on Life Cycle Skill Formation," Discussion Paper NBER Working Paper No. 11331.

Dehejia, R., and S. Wahba (1999): "Causal Effects in Nonexperimental Studies: Re-Evaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053–1062.

Froelich, M. (2004): "Matching Estimators and Optimal Bandwidth Choice," Discussion paper, University of St. Gallen.

Garen, J. (1984): "The Returns to Schooling: A Selectivity Bias Approach with a Continous Choice Variable," *Econometrica*, 52, 1199 – 1218.

Heckman, J., and V. Hotz (1989): "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs," *Journal of the American Statistical Association*, 84, 862–874.

Heckman, J. J., J. H. Ichimura, J. Smith, and P. Todd (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017 – 1098.

Heckman, J. J., J. H. Ichimura, and P. Todd (1997): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261 – 294.

Heckman, J. J., and E. Vytlacil (1998): "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Rate of Return to Schooling When the return is Corelated with Schooling," *Journal of Human Resources*, 23, 974–987.

Imbens, G. (2003): "Nonparametric Estimation of Average Treatment Effects Under Exogeneity," Discussion paper, NBER Technical Working Paper 294.

Juhn, C., K. Murphy, and B. Pierce (1993): "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, 101, 410 – 442.

Katz, L., and D. Autor (1999): "Changes in the Wage Structure and Earnings Inequality," in *Handbook of Labor Economics, Vol. 3A*, ed. by O. Ashenfelter, and D. Card, pp. 1463 – 1555. Elsevier Science, Amsterdam.

LaLonde, R. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76, 604–620.

Mincer, J. (1974): *Schooling, Experience and Earnings*. Columbia University Press, New York.

Smith, J., and P. Todd (2001): "Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods," *American Economic Review*, 91, 112–118.

——— (2005): "Does Matching Overcome LaLondes Critique of Nonexperimental Estimators?," *Journal of Econometrics*, 125, 305–353.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

——— (2004): "Estimating Average Partial Effects under Conditional Independence Assumptions," unpublished working paper, Department of Economics, Michigan State University.

Appendix

Proof of Proposition 2.1

Take the expectation of (2.1) conditional on α, β and X and subtract it from the original equation:

$$\begin{aligned} \mathrm{E}\left[Y|\alpha,\beta,X\right] &= \alpha + \beta\,\mathrm{E}\left[S|\alpha,\beta,X\right] \\ (Y - \mathrm{E}\left[Y|\alpha,\beta,X\right]) &= \beta \cdot (S - \mathrm{E}\left[S|\alpha,\beta,X\right]) \end{aligned}$$

Multiply both sides of the equation by $(S - \mathrm{E}\left[S|\alpha,\beta,X\right])$ to get

$$(S - \mathrm{E}\left[S|\alpha,\beta,X\right])(Y - \mathrm{E}\left[Y|\alpha,\beta,X\right]) = \beta(S - \mathrm{E}\left[S|\alpha,\beta,X\right])^2.$$

The expectation of both sides of the equation on α, β and X is:

$$\begin{aligned} \mathrm{Cov}\left[S,Y|\alpha,\beta,X\right] &= \beta\,\mathrm{V}\left[S|\,\alpha,\beta,X\right] \\ &= \beta\,\mathrm{V}\left[S|\,X\right], \qquad (.1) \end{aligned}$$

where the rhs of the second equality results from ignorability assumption *iii b)*. Note that that under ignorability conditions *ii)* and *iii a)*:

$$\begin{aligned} & \mathrm{E}\left[\mathrm{Cov}\left[S,Y|\alpha,\beta,X\right]|\,X\right] \\ = \;& \mathop{\mathrm{E}}_{\alpha,\beta}\left[\mathrm{E}\left[YS|\,\alpha,\beta,X\right]|X\right] - \mathop{\mathrm{E}}_{\alpha,\beta}\left[\mathrm{E}\left[Y|\,\alpha,\beta,X\right]\mathrm{E}\left[S|\,\alpha,\beta,X\right]|X\right] \\ = \;& \mathrm{E}\left[YS|\,X\right] - \mathop{\mathrm{E}}_{\alpha,\beta}\left[\mathrm{E}\left[Y|\,\alpha,\beta,X\right]\mathrm{E}\left[S|\,X\right]|X\right] \\ = \;& \mathrm{E}\left[YS|\,X\right] - \mathrm{E}\left[Y|\,X\right]\mathrm{E}\left[S|\,X\right] \\ = \;& \mathrm{Cov}\left[S,Y|X\right]. \end{aligned}$$

Taking the expectation on both sides of the equation (.1) over α and β conditional on X:

$$\mathrm{Cov}\left[S,Y|X\right] = \mathrm{E}\left[\beta|\,X\right]\mathrm{V}\left[S|\,X\right]. \qquad (.2)$$

Solving for $\mathrm{E}\left[\beta|X\right]$ and applying the law of iterated expectations gives the desired result.

$\square$

Proof of Lemma 2.1

By the law of iterated expectations and ignorability condition *iii a)*:

$$\begin{aligned}
\mathrm{E}\left[\beta S|X\right] &= \mathop{\mathrm{E}}_{\alpha,\beta}\left[\mathrm{E}\left[\beta S|\,\alpha,\beta,X\right]|X\right] \\
&= \mathop{\mathrm{E}}_{\alpha,\beta}\left[\beta\,\mathrm{E}\left[S|\,\alpha,\beta,X\right]|X\right] \\
&= \mathop{\mathrm{E}}_{\alpha,\beta}\left[\beta\,\mathrm{E}\left[S|\,X\right]|X\right] \\
&= \mathrm{E}\left[\beta|\,X\right]\cdot\mathrm{E}\left[S|\,X\right]
\end{aligned}$$

Given the uncorrelatedness $\mathrm{Cov}\left[\beta, S|X\right] = \mathrm{E}\left[\beta S|X\right] - \mathrm{E}\left[\beta|\,X\right]\cdot\mathrm{E}\left[S|\,X\right]$ we obtain for α:

$$\begin{aligned}
\mathrm{E}\left[\alpha|X\right] &= \mathrm{E}\left[Y|X\right] - \mathrm{E}\left[\beta S|\,X\right] \\
&= \mathrm{E}\left[Y|X\right] - \mathrm{E}\left[\beta|\,X\right]\cdot\mathrm{E}\left[S|\,X\right].
\end{aligned}$$

Since all tree terms on the rhs are identified, the unconditional mean is identified by interpretation of X

$$\begin{aligned}
\mathrm{E}\left[\alpha\right] = \mathrm{E}\left[\mathrm{E}\left[\alpha|\,X\right]\right] &= \mathrm{E}\left[Y\right] - \mathrm{E}\left[\mathrm{E}\left[\beta|\,X\right]\cdot\mathrm{E}\left[S|\,X\right]\right] \\
&= \mathrm{E}\left[Y\right] - \mathrm{E}\left[\frac{\mathrm{Cov}\left[S,Y|X\right]\cdot\mathrm{E}\left[S|X\right]}{\mathrm{V}[S|X]}\right].
\end{aligned}$$

□

Proof of Proposition 2.2

Using the assumptions and result of Proposition 2.1 only $\mathrm{E}\left[\beta^2\right]$ needs to be identified to identify $\mathrm{V}\left[\beta\right] = \mathrm{E}\left[\beta^2\right] - \mathrm{E}\left[\beta\right]^2$.

The conditional expectation $\mathrm{E}\left[\beta^2|\,X\right]$ can be obtained similarly to the computation of $\mathrm{E}\left[\beta|\,X\right]$. Take the expectation of equation (2.1) conditional on α, β and X and subtract it from the original equation which gives after squaring:

$$(Y - \mathrm{E}\left[Y|\alpha,\beta,X\right])^2 \;=\; \beta^2(S - \mathrm{E}\left[S|\alpha,\beta,X\right])^2.$$

The expectation of both sides of the equation on α, β and X is:

$$\mathrm{V}\left[Y|\,\alpha,\beta,X\right] = \beta^2\,\mathrm{V}\left[S|\,X\right].$$

Taking the expectation on both sides conditional on X and solving for $\mathrm{E}\left[\beta^2 \middle| X\right]$ gives:

$$\mathrm{E}\left[\beta^2 \middle| X\right] = \frac{\mathrm{E}_{\alpha,\beta}\left[\mathrm{V}\left[Y|\alpha,\beta,X\right]\right]}{\mathrm{V}\left[S|X\right]} \tag{.3}$$

The term in the numerator is given by

$$\begin{aligned}\underset{\alpha,\beta}{\mathrm{E}}\left[\mathrm{V}\left[Y|\,\alpha,\beta,X\right]\right] &= \mathrm{V}\left[Y|X\right] - \underset{\alpha,\beta}{\mathrm{V}}\left[\mathrm{E}\left[Y|\,\alpha,\beta,X\right]|X\right] \\ &= \mathrm{V}\left[Y|\,X\right] - \mathrm{V}\left[\alpha|\,X\right] - \mathrm{V}\left[\beta|\,X\right]\mathrm{E}\left[S|\,X\right]^2 - 2\mathrm{Cov}\left[\alpha,\beta|X\right]\mathrm{E}\left[S|\,X\right].\end{aligned}$$

Since $\mathrm{V}\left[\alpha|\,X\right]$ is identified by the ignorability conditions $\mathrm{E}\left[\beta^2|\,X\right]$ is identified if $\mathrm{Cov}\left[\alpha,\beta|X\right]$ is identified.

□

Proof of Proposition 2.3

Consider the numerator of .3 and assume $\mathrm{Cov}\left[\alpha,\beta|X\right] \leq 0$. Then,

$$\mathrm{E}\left[\beta^2 \middle| X\right] \geq \frac{\mathrm{V}\left[Y|\,X\right] - \mathrm{V}\left[\alpha|\,X\right] - \mathrm{E}\left[S|\,X\right]^2\mathrm{V}\left[\beta|\,X\right]}{\mathrm{V}\left[S|\,X\right]}.$$

Solving for $\mathrm{E}\left[\beta^2|\,X\right]$ results in:

$$\mathrm{E}\left[\beta^2 \middle| X\right] \geq \frac{\mathrm{V}\left[Y|\,X\right] - \mathrm{V}\left[\alpha|\,X\right] + \mathrm{E}\left[\beta|\,X\right]^2\mathrm{E}\left[S|\,X\right]^2}{\mathrm{E}\left[S^2|\,X\right]}.$$

□

Proof of Proposition 2.4

Using the variance decomposition for the numerator of .3

$$\underset{\alpha,\beta}{\mathrm{E}}\left[\mathrm{V}\left[Y|\alpha,\beta,X\right]\right] = \mathrm{V}\left[Y|X\right] - \underset{\alpha,\beta}{\mathrm{V}}\left[\mathrm{E}\left[Y|\,\alpha,\beta,X\right]|X\right],$$

and dropping the unidentified second term of the difference we obtain an upper bound for $\mathrm{E}\left[\beta^2\right]$:

$$\mathrm{E}\left[\beta^2\right] = \underset{X}{\mathrm{E}}\left[\mathrm{E}\left[\beta^2 \middle| X\right]\right] \leq \underset{X}{\mathrm{E}}\left[\frac{\mathrm{V}\left[Y|X\right]}{\mathrm{V}\left[S|X\right]}\right]$$

□