ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Universität
Konstanz
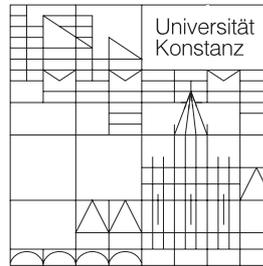
**Rechts-, Wirtschafts- und
Verwaltungswissenschaftliche
Sektion**
Fachbereich
Wirtschaftswissenschaften

Diskussionspapiere der DFG-
Forschergruppe (Nr.: 3468269275):

Heterogene Arbeit: Positive und Normative
Aspekte der Qualifikationsstruktur der Arbeit

Markus Jochmann,
Winfried Pohlmeier

**The Causal Effect of Overqualification on
Earnings: Evidence from a Bayesian
Approach**

Februar 2004

The Causal Effect of Overqualification on Earnings: Evidence from a Bayesian Approach

Markus Jochmann

Fach D124
78457 Konstanz
Germany

mail: markus.jochmann@uni-konstanz.de

Winfried Pohlmeier

Fach D 124
78457 Konstanz
Germany

mail   : winfried.pohlmeier@uni-konstanz.de
phone : +49-7531-88-2660
fax   : +49-7531-88-4450

Zusammenfassung:

This paper focuses on the causal effect of overqualification on earnings. Although the issue of overqualification has recently been addressed by quite a huge body of literature there are only few studies examining the causal effect of overqualification on earnings in the sense of Rubin's potential outcome approach. Since for non-experimental data settings the incident of overqualification is not a random event, ignoring selfselection into overqualification leads to a misinterpretation of the empirical results and misleading policy conclusions, since the effect of overqualification on earnings cannot be interpreted causally.
Using a cross-section of 1188 workers from the GSOEP we apply a Bayesian approach based on Markov chain Monte Carlo methods to estimate various treatment effects of overqualification on earnings. Our findings seriously question results ignoring selectivity effects and point out that on average for the overeducated an appropriate job match would not lead to higher earnings.

The Causal Effect of
Overqualification on Earnings:
Evidence from a Bayesian Approach

Markus Jochmann[*]

University of Konstanz

Winfried Pohlmeier[†]

University of Konstanz, CoFE, ZEW

Abstract

This paper focuses on the causal effect of overqualification on earnings. Although the issue of overqualification has recently been addressed by quite a huge body of literature there are only few studies examining the causal effect of overqualification on earnings in the sense of Rubin's potential outcome approach. Since for non-experimental data settings the incident of overqualification is not a random event, ignoring selfselection into overqualification leads to a misinterpretation of the empirical results and misleading policy conclusions, since the effect of overqualification on earnings cannot be interpreted causally.

Using a cross-section of 1188 workers from the GSOEP we apply a Bayesian approach based on Markov chain Monte Carlo methods to estimate various treatment effects of overqualification on earnings. Our findings seriously question results ignoring selectivity effects and point out that on average for the overeducated an appropriate job match would not lead to higher earnings.

JEL classification: $C15, C24, J24, J31$

Keywords: returns to schooling, overqualification, potential outcome approach, latent index selection model, Markov chain Monte Carlo

1 Introduction

Among economists overqualification is widely regarded as one important dimension of measuring the inefficiency of a labor market. The extent of the inefficient use of labor through overqualification is, however, hard to measure. The common approach so far has been to compare the returns of overqualification with the returns of required schooling based on an augmented Mincer-type earnings function.[1] Such an approach provides only limited insight into the degree of the inefficiencies due to overqualification. There are two distinct reasons for this. First, even if the returns of overqualification are lower than the returns of required schooling, as the international evidence suggests, lower return rates do not necessarily imply inefficient use of labor and/or suboptimal behavior of workers. This is because lower return rates of overqualification might simply be seen as evidence for decreasing returns of schooling. Secondly, the negative wage effects of overqualification may be due to nonrandom selection into overqualification, that is a position for which one is overqualified. This situation induces spurious correlations between overqualification and earnings. In fact, ignoring selfselection into overqualification leads to a misinterpretation of the empirical results and consequently to misleading policy conclusions. The effect of overqualification on earnings cannot be interpreted causally.[2]

In this paper, we estimate the causal effect of overqualification on earnings using the potential outcomes approach due to Rubin (1974). This approach has been adopted in numerous econometric evaluation studies, including those of Angrist, Imbens, and Rubin (1996), Heckman (1992, 1997) and Heckman, Ichimura, and Todd (1997). Specifically, we focus on the estimation of the average treatment on the treated effect (TT) of overqualification on earnings, which provides us with a measure for the income loss for the overqualified as caused by overqualification. Basically, there are two alternative approaches to identifying the counterfactual outcome needed to compute this treatment effect. The basic identification condition of the matching approaches is the conditional mean independence assumption introduced by Rosenbaum and Rubin (1983). This identification condition is particularly reasonable if sufficient observable covariates are available to control for the selection process ("se-

---

[1]See e.g. Groot and Maassen van den Brink (2000) for a comprehensive survey on the international evidence.

[2]Only a few studies raise the issue of selectivity in the context of overqualification, recent exceptions are Bauer (2002) and Büchel and van Ham (2003).

lection on observables"). In this study, we adopt another approach that models the selection into the treatment and the non-treatment group explicitly by means of a latent index selection equation. This equation includes observables (instruments) and unobservables ("selection on unobservables").[3] Despite the obvious advantages of structural approaches (e.g. parameters and identifying restrictions, which can easily be interpreted in economic terms, out of sample forecasting, etc.), a major drawback of this approach is the difficulty of finding plausible exclusion restrictions (instruments). Moreover, instruments, although economically reasonable, might be weak which questions the consistency and asymptotic normality of the estimator that only holds true in the case of strong instruments (see Staiger and Stock (1997)).

Here, we apply a Bayesian analysis based on Markov chain Monte Carlo (MCMC) methods in order to analyze the latent index selection model which yields small sample evidence on the average treatment effect and the average treatment on the treated effect. Moreover, the implementation of MCMC techniques is connected with further advantages as discussed below.

The sequel of the paper is organized as follows. In Section 2, we set up the standard sample selection approach for evaluating the binary treatment effects. Then we relate the treatment parameters to the selection model and to approaches ignoring the selectivity problem. Our database is briefly described in Section 3. In Section 4, we briefly discuss the Bayesian approach to the estimation of treatment effects following the methods by Chib and Hamilton (2000) and present the empirical findings. In particular, we determine which type of worker is affected most by income losses. Section 5 concludes and provides an outlook on future research.

2   The Model

Following the treatment effect literature (see for example, Fisher (1935), Roy (1951) and Cox (1958)), we specify two potential outcomes (Y_{0i}, Y_{1i}) for each person i. The corresponding outcome equations are modeled as:

$$Y_{0i} = X_i\beta_0 + U_{0i}, \tag{2.1}$$

---

[3]Vytlacil (2002) shows that the nonparametric version of the latent index selection model and the instrumental variable approaches estimating the local average treatment effect (LATE) proposed by Imbens and Angrist (1994) use equivalent identifying restrictions.

respectively

$$Y_{1i} = X_i\beta_1 + U_{1i}, \tag{2.2}$$

where X_i is a vector of observed random variables and U_{0i} and U_{1i}, respectively, are unobserved random variables. In our case, Y_{0i} denotes the log earnings of a person if that person has an adequate job, and Y_{1i} measures the log earnings of that person in the case of overqualification. The vector X_i contains the usual control variables as age, experience, etc. In this setup, the difference

$$\Delta_i = Y_{1i} - Y_{0i} \tag{2.3}$$

can be interpreted as the causal effect of overqualification on earnings. The basic problem now is that we cannot observe both Y_{0i} and Y_{1i} for the same individual. Instead, we observe Y_{0i} for the adequately qualified only and Y_{1i} only for overqualified persons:

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}, \tag{2.4}$$

where $D_i = 1$ denotes overqualification and $D_i = 0$ denotes adequate qualification. The individual decision rule is finally specified by

$$D_i^* = W_i\beta_D + U_{Di}, \tag{2.5}$$

$$D_i = \begin{cases} 1, & \text{if } D_i^* \geq 0, \\ 0, & \text{otherwise}, \end{cases} \tag{2.6}$$

where D_i^* is a latent variable, determining whether the individual is overeducated or not, and W_i a vector of observed random variables.

Though the individual treatment effect Δ_i cannot be calculated, we can determine the average treatment effect (ATE) and the average effect of the treatment on the treated (TT) for the subpopulation with $X_i = x$, which are given by:

$$\begin{aligned} \Delta^{\text{ATE}}(x) &= \text{E}\left[\Delta_i | X_i = x\right] \\ &= \text{E}\left[Y_{1i} | X_i = x\right] - \text{E}\left[Y_{0i} | X_i = x\right], \end{aligned} \tag{2.7}$$

$$\begin{aligned} \Delta^{\text{TT}}(x) &= \text{E}\left[\Delta_i | D_i = 1, X_i = x\right] \\ &= \text{E}\left[Y_{1i} | D_i = 1, X_i = x\right] - \text{E}\left[Y_{0i} | D_i = 1, X_i = x\right]. \end{aligned} \tag{2.8}$$

In general, these two treatment effects are different due to the unobserved errors U_{0i} and U_{1i}. However, in the case that the difference Δ_i is the same for all individuals with characteristics x the two treatment effects coincide.[4] This is also true in the situation where there are idiosyncratic gains from the treatment, but individuals do not take these into account while deciding whether to participate in the program or not, see Heckman (1997).

In this paper, we focus on unconditional estimates that we obtain by integrating equation (2.7) over the distribution of X_i for all individuals and respectively, by integrating euqation (2.8) over the joint distribution of X_i and W_i for those who are in the treatment state:

$$\Delta^{\text{ATE}} = \int \Delta^{\text{ATE}}(x) \, dF(x), \tag{2.9}$$

$$\Delta^{\text{TT}} = \int \Delta^{\text{TT}}(x) \, dF(x|D = 1). \tag{2.10}$$

To close the model we follow a suggestion by Aakvik, Heckman, and Vytlacil (2000) and impose the following factor structure with one unobservable factor θ_i on the error terms:

$$U_{0i} = \alpha_0 \theta_i + \varepsilon_{0i}, \tag{2.11}$$

$$U_{1i} = \alpha_1 \theta_i + \varepsilon_{1i}, \tag{2.12}$$

$$U_{Di} = \alpha_D \theta_i + \varepsilon_{Di}, \tag{2.13}$$

where ε_{0i}, ε_{1i}, ε_{Di} and θ_i are independently normally distributed with mean zero and are independent of X_i and W_i. In order to identify the model we finally assume $\alpha_D = 1$ and $\text{V}[\theta_i] = \text{V}[\varepsilon_{Di}] = 1$. By assuming joint normality of the error terms, this assumption completely identifies the joint distribution of U_{Di}, U_{0i} and U_{1i}.[5]

---

[4]The resulting model is then called the *dummy endogenous variable model*.
[5]α_0 is identified by $\text{Cov}[U_{Di}, U_{0i}]$, α_1 by $\text{Cov}[U_D, U_1]$. Thus, $\text{Cov}[U_{0i}, U_{1i}] = \alpha_0 \alpha_1$ is identified.

3 The Data

In our empirical analysis we use data from the 1998 wave of the German Socio Economic Panel (GSOEP). The GSOEP is a representative longitudinal survey of German households conducted by the German Institute for Economic Research in Berlin.[6]

In the literature three main strategies for measuring overqualification are found: systematic job analysis (the so called "objective approach", where job analysts determine the required level of education), self-assessment by workers (the so called "subjective approach", where the worker himself specifies the education that he needs to perform his job) and the analysis of realized matches (here the required level of education is derived from the average worker's education in the considered job).[7] In this study we apply a variant of the 'subjective approach' that was proposed by Büchel and Weisshuhn (1998). They use the answer to the GSOEP question: "What type of training is normally required for the job you do?" in order to determine the level of congruence between the actual and required education. Then they validate the result considering the occupational status of the individual. In this way, they are able to classify workers into five categories. In this study we compare individuals who are "(definitely) overqualified" with those who fall into the category with "(definite) congruence between job and education".

Furthermore, our empirical analysis is restricted to full-time employed, prime-aged (18-60 years) men of German nationality who obtained a school leaving certificate in West Germany. In addition, self-employed and those currently in education or training are excluded. After eliminating all observations with missing values, we obtain a final sample of 1188 individuals with 141 of them classified as overqualified. The variable definitions and summary statistics are reported in Table 1.

---

[6]For more information, see SOEP Group (2001).

[7]For a detailed analysis of these three strategies see Chapter 1 of this book.

Table 1: Variable Definitions and Summary Statistics

Variable	Description	Mean	Std. Dev.
Overqual	Dummy for overqualification		
	(1 = overqualified, 0 = adequately qualified)	0.119	
Logearn	Log of gross monthly earnings[a]	8.648	0.425
Partner	Dummy for family status		
	(1 = married or lives with partner, 0 = otherwise)	0.798	
Age	Age (years / 10)	3.817	0.824
Age2	Age squared (years / 100)	15.249	6.391
Firm20	Dummy for firm size		
	(1 = 20 ≤ Firm size < 2000, 0 = otherwise)	0.515	
Firm2000	Dummy for firm size		
	(1 = 2000 ≤ Firm size, 0 = otherwise)	0.310	
Civil	Dummy for civil service		
	(1 = individual is civil servant, 0 = otherwise)	0.238	
Tenure	Firm tenure (years / 10)	1.098	0.901
Health	Dummy for health		
	(1 = individual is restricted, 0 = otherwise)[b]	0.181	
Education	Years of education	12.626	2.588
UR	Regional unemployment rate[c]	4.315	3.381

[a] Full time equivalent based on 40 working hours per week, in DM
[b] Person is restricted on account of health reasons in carrying out daily tasks
[c] At time of first school graduation in the respective federal state

Table 2 contains the summary statistics of the covariates for the overqualified and adequately qualified individuals separately. The test of equality of the group means (last column) does not indicate significant differences for most of the covariates. The overqualified workers can be found more often in the private sector and in smaller firms. Mean tenure for the overqualified is lower than for the adequately employed workers. Note, that mean schooling for the overqualified is smaller than for the adequately employed.

Table 2: Summary Statistics for the Two Groups

Variable	Adequately employed workers		Overqualified workers		p-value
	Mean	Std. Dev.	Mean	Std. Dev.	
Logearn	8.681	0.421	8.405	0.369	0.000
Partner	0.802		0.766		0.338
Age	3.830	0.826	3.721	0.808	0.133
Age2	15.352	6.428	14.491	6.076	0.119
Firm20	0.508		0.567		0.186
Firm2000	0.318		0.248		0.077
Civil	0.250		0.149		0.002
Tenure	1.150	0.914	0.708	0.678	0.000
Health	0.178		0.206		0.439
Education	12.762	2.626	11.617	2.012	0.000
UR	4.290	3.351	4.495	3.604	0.524

Figure 1 shows kernel density estimates of the log earnings for the two groups. Interestingly, in terms of income variation the group of the overqualified seems to be less heterogenous than the adequately employed group. Adequate employment goes along with higher average earnings but also with a higher earnings variation.
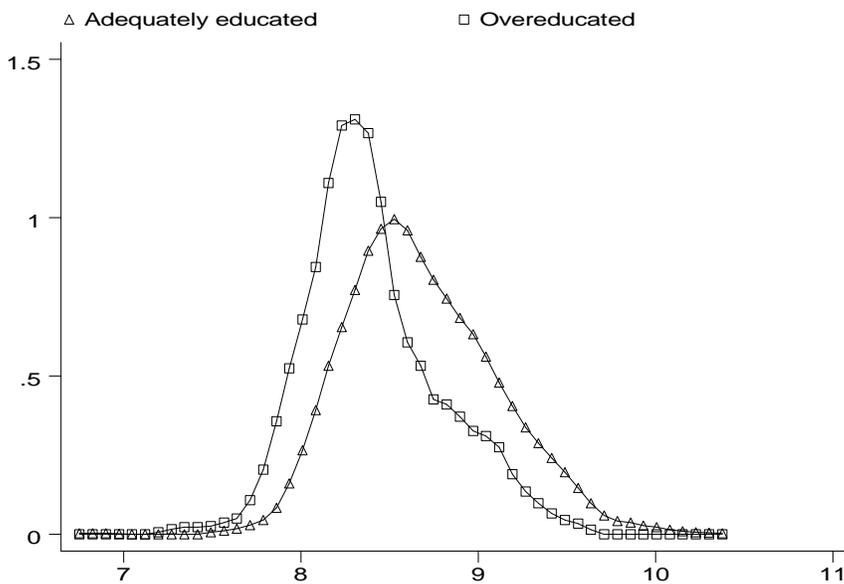
Figure 1: Kernel density estimates of log earnings

4   Results

We estimate the model using Markov chain Monte Carlo (MCMC) techniques. This means that we base our inference on a large sample of draws from the posterior distribution, where the sample is generated by designing a Markov chain with a transition kernel having an invariant measure equal to the posterior distribution.[8] Furthermore we augment the parameter space by including the latent variables (following Tanner and Wong (1987)).

We chose the prior distributions to express prior ignorance. We ran the MCMC algorithm for 120.000 iterations keeping every 10th of the the last 100.000 iterations. The mixing behavior of the chain was satisfying. A detailed description of the sampling scheme and the prior distributions used is given in the appendix.

The Bayesian approach and its application via MCMC methods has several advantages. First, we can easily compute subject-specific treatment distributions by simulating the latent variables (see Chib and Hamilton (2000)). We do not report these here, but they are a great help in calculating the mean treatment effects discussed in this paper. Second, we do not have to apply numerical integration methods to evaluate the likelihood function of our model. This is of particular importance when more flexible distributions than the normal are assumed for the error terms. Finally, the Bayesian approach allows us to conduct exact small sample inference. Given the rather small number of overqualified in our data, the reliability on large sample approximations is questionable.

Although the factor structure of the error terms and the normality assumption completely identify the latent index selection model we use the regional unemployment rate at graduation and its interaction term with age as additional instruments in the selection equation following the reasoning put forward by Maier, Pfeiffer, and Pohlmeier (2003). They argue that times with inferior employment prospects cause students to extend their participation in the educational sector due to lower opportunity costs. In particular, the preparation year for vocational training (Berufsvorbereitungsjahr, BVJ) and the elementary vocational year (Berufsgrundbildungsjahr,

---

[8]See Chen, Shao, and Ibrahim (2000) or Robert and Casella (1999) for details on MCMC methods.

BGJ) are compulsory for youths who have not received an offer for an apprentice-ship training position and are below the compulsory schooling age of eighteen. The two institutions, therefore, serve to circumvent youth unemployment.[9] However, the argument also holds without explicit reference to the institutional setup of the BVJ and the BGJ and may also be valid for university graduates. Given that unemployment reflects opportunity costs, an individual is more likely to stay in the educational system if employment prospects are low. Clearly, this argument seems particularly relevant for the case of Germany where tuition and fees for general schooling and vocational training are rare exceptions or negligible.

Table 3 contains the estimation results for the selection equation. The findings are largely in accordance with the results of previous studies. The probability of being overqualified is lower for workers with a higher educational attainment. The bell-shaped effect of age on the probability of overqualification is difficult to interpret, since our cross-section data do not allow for discrimination between age, cohort and time effects. For workers born before 1961 ($Age \geq 37$), the probability of overqualification decreases with age. A positive effect of age on the probability of overqualification can be found for younger cohorts ($Age < 37$). The coefficient estimates for the unemployment rate at the time of graduation and the interaction term with Age are quite precise. The sign of the effect of unemployment on the probability of overqualification is ambiguous due to the negative coefficient of the interaction term. Thus, the effect of unemployment on the probability of being overqualified is positive for younger cohorts and negative for the older ones. A positive relationship between overqualification and unemployment can be found for workers who are younger than 30 years. This will be due to the fact that the unemployment rate at the time of graduation proxies the actual unemployment rate for the younger age groups. This means that unemployment and overqualification express excess supply of skilled workers, see Velling and Pfeiffer (1997) who find a positive correlation between the actual unemployment rate and the incidence of overqualification.

---

[9]Basically the BVJ and the BGJ serve the same purpose. The BVJ is nowadays quantitatively far more important than the BGJ. The latter can be found in larger vocational schools and often serves as a substitute for the first year within the dual vocational training system. Franz, Inkmann, Pohlmeier, and Zimmermann (2000) give a brief description of the German vocational training system.

Table 3: Estimates of the selection equation

	Mean	Std. Dev.
Constant	-6.368	3.349
Partner	-0.099	0.181
Age	3.553	1.618
Age2	-0.387	0.191
Firm20	0.276	0.195
Firm2000	0.308	0.227
Civil	-0.188	0.193
Tenure	-0.769	0.116
Health	0.104	0.181
Edu	-0.183	0.035
UR	0.468	0.220
UR $\times$ Age	-0.154	0.065

Means and standard deviations of the posterior distributions

In order to work out the robustness of earlier empirical findings that neglect sample selectivity, we present estimation results for two model specifications in the following section. One specification disregards non-random selectivity while the other takes selectivity into account through the selection equation. Table 4 reports on the coefficients of the earnings functions for overqualified and adequately qualified workers. The posterior means of the correlation coefficients between the earnings equations and the selection equation are -0.294 and 0.260 respectively, but are not well concentrated around these values. Accounting for selectivity does not seriously affect the sign pattern of the posterior means. Note, however, that the parameters of the specification with selectivity are estimated with much greater precision than the model ignoring selectivity.

As expected, the parameters of the earnings equations for the overqualified are not as well determined as they are for adequately employed workers. The returns to schooling for the adequately employed are around 7%, which is very similar to estimates for Germany obtained by classical methods. Overqualified workers clearly have to face lower return rates, which is consistent with previous empirical evidence (see Groot and Maassen van den Brink (2000)).

Table 4: Estimates of the earnings equations

	Model without selection		Model with selection	
	Overqual=0	Overqual=1	Overqual=0	Overqual=1
Constant	6.360	7.724	6.300	7.706
	(1.479)	(2.340)	(0.226)	(0.690)
Partner	0.079	0.174	0.085	0.159
	(0.192)	(0.264)	(0.029)	(0.078)
Age	0.563	0.112	0.556	0.106
	(0.782)	(1.272)	(0.116)	(0.374)
Age2	−0.053	−0.016	−0.053	−0.012
	(0.101)	(0.167)	(0.015)	(0.049)
Firm20	0.112	−0.007	0.103	0.010
	(0.204)	(0.304)	(0.030)	(0.092)
Firm2000	0.182	0.052	0.170	0.072
	(0.226)	(0.336)	(0.033)	(0.104)
Civil	−0.167	−0.072	−0.165	−0.067
	(0.176)	(0.314)	(0.026)	(0.093)
Tenure	−0.016	0.039	0.001	−0.002
	(0.110)	(0.179)	(0.022)	(0.080)
Health	−0.068	0.142	−0.074	0.143
	(0.187)	(0.273)	(0.028)	(0.082)
Education	0.068	0.026	0.072	0.014
	(0.030)	(0.053)	(0.006)	(0.023)
Corr			−0.294	0.260
			(0.290)	(0.342)
σ^2	0.171	0.386	0.075	0.094
	(0.024)	(0.046)	(0.036)	(0.041)

Means and standard deviations of the posterior distributions

As argued above, a comparison of the coefficients on the schooling variable in the two earnings equations provides only limited information on the causal effect of overqualification on earnings. We, therefore, present the Bayesian estimates of the average treatment effect (ATE) and the average treatment on the treated effect (TT) in Table 5. According to the definitions for the two treatment effects given above,

a negative TT effect indicates that an overqualified worker, if employed adequately, can expect an increase in earnings. A negative ATE means that on average, the earnings from adequate employment are larger than the earnings from a job where the worker is overqualified. This holds true when workers are randomly selected into the two types of jobs. Using the model that accounts for selectivity, the mean of the posterior distribution of the TT is close to zero with a large standard deviation of the posterior. That is, our results do not indicate that overqualified workers can expect significant earnings increases if they are placed in jobs where their educational backgrounds are adequate.

Table 5: Estimates of the ATE and the TT

	Model without selection	Model with selection
ATE	-0.230	-0.395
	(0.129)	(0.262)
TT	-0.193	-0.007
	(0.094)	(0.187)

Means and standard deviations of the posterior distributions

The kernel density estimates of the posterior distributions of the TT for the two specifications depicted in Figure 2 clarify that our results are driven by the selection on unobservables. Comparing the TT effects based on the two alternative specification shows that accounting for non-random selectivity leads to an imprecise estimate of the TT effect.
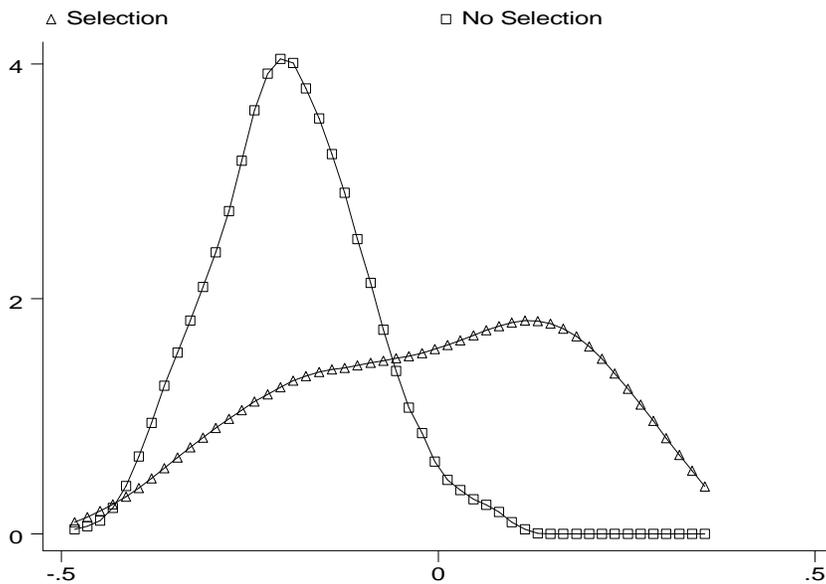
Figure 2: Kernel density estimates of the TT effect

Based on the estimates presented above we conclude with an analysis of the causal effect of overqualification on earnings for various subgroups of the population. This is done to determine which type of workers is affected most by income losses due to inadequate employment. We differentiate types of workers according to their backgrounds in general schooling (lower secondary school, secondary school and Gymnasium) and their ages by dividing them into three different age groups. The results for the average treatment on the treated effect are given in Table 6. The negative earnings impact of overqualification is more distinct for workers with a higher level of education and for older workers (but note the high standard deviations).

Table 6: Estimates of the TT for selected groups

	Model without selection	Model with selection
Lower secondary school	−0.174	0.005
	(0.106)	(0.181)
Secondary school	−0.215	−0.022
	(0.113)	(0.198)
Gymnasium	−0.330	−0.090
	(0.209)	(0.265)
Age < 30	−0.005	0.175
	(0.175)	(0.189)
30 ≤ Age < 50	−0.240	−0.053
	(0.103)	(0.189)
50 ≤ Age	−0.284	−0.089
	(0.240)	(0.216)

Means and standard deviations of the posterior distributions

5 Conclusions

This chapter provides empirical evidence of the causal effect of overqualification on earnings in the vein of Rubin (1974). Using a Bayesian approach, we estimated the average treatment effect and the treatment on the treated effect on earnings due to overqualification based on a latent index selection model. Based of our results, we argue that the TT can serve as a benchmark for measuring labor market inefficiency caused by overqualification. Contrary to previous studies neglecting selectivity issues, we find no evidence supporting the idea that overqualification (on average) depresses earnings. If overqualified workers were placed in jobs where the job requirements were in accordance with their actual educational attainment, they could not expect significant increases in earnings.

Although an overqualified worker cannot necessarily expect an income gain if adequately employed, our findings do not imply that overqualification does not occur without any cost from the social welfare point of view. Obviously, earnings studies such as this do not take into account individual costs (monetary and time costs) of

overqualification or the social costs of overqualification. Moreover, focusing on mean income differences between overeducated and adequately employed workers ignores the possibility of income risk being a determinant of job choice, i.e., a mismatch between job requirements and the educational attainment of the worker may result in utility losses due to a higher job risk.

The Bayesian approach presented here proves to be attractive from theoretical and practical points of view. The possibility to estimate treatment effects on the individual level and exact small sample inference are just two advantages. Moreover, the coefficient estimates based on an uninformative prior are very similar to the results obtained by classical procedures. But computing the whole distribution of the potential outcomes creates an opportunity for analyzing a range of treatment effects within the potential outcome approach. An obvious caveat of the analysis presented here is the lack of evidence for the robustness of our results. Future research using the Bayesian approach should be concerned with alternative distributional assumptions of the error terms and the robustness of the results with respect to the identifying restrictions.

References

AAKVIK, A., J. J. HECKMAN, AND E. J. VYTLACIL (2000): "Treatment Effects for Discrete Outcomes when Responses to Treatment Vary Among Observationally Identical Persons: An Application to Norwegian Vocational Rehabilitation Programs," NBER Technical Working Paper, No. 262.

ANGRIST, J., G. W. IMBENS, AND D. B. RUBIN (1996): "Identification of Causal Effects Using Instrumental Variables (with discussion)," *Journal of the American Statistical Association*, 91, 444 – 472.

BAUER, T. (2002): "Educational Mismatch and Wages in Germany," *Economics of Education Review*, 21, 221–229.

BÜCHEL, F. (2001): "Overqualification: Reasons, Measurement Issues and Topological Affinity to Unemployment," in *Training in Europe. Second Report on Vocational Training Research in Europe 2000: Background Report*, ed. by P. Descy, and M. Tessaring, vol. 2, pp. 453–560. CEDEFOP Reference Series, Luxembourg.

BÜCHEL, F., AND M. VAN HAM (2003): "Overeducation, Regional Labour Markets and Spatial Flexibility," forthcoming, Journal of Urban Economics.

BÜCHEL, F., AND G. WEISSHUHN (1998): *Ausbildungsinadäquate Beschäftigung der Absolventen des Bildungssystems - II. Fortsetzung der Berichterstattung zu Struktur und Entwicklung unterwertiger Beschäftigung in West- und Ostdeutschland (1993- 1995)*. Report on Behalf of the Federal Ministry of Education and Research.

CHEN, M. H., Q. M. SHAO, AND J. G. IBRAHIM (2000): *Monte Carlo Methods in Bayesian Computation*. Springer Verlag, New York.

CHIB, S., AND B. H. HAMILTON (2000): "Bayesian Analysis of Cross-Section and Clustered Data Treatment Models," *Journal of Econometrics*, 97, 25 – 50.

COX, D. R. (1958): *The Planning of Experiments*. Wiley, New York.

FISHER, R. A. (1935): *Design of Experiments*. Oliver and Boyd, London.

FRANZ, W., J. INKMANN, W. POHLMEIER, AND V. ZIMMERMANN (2000): "Young and Out in Germany: On Youths' Chances of Labor Market Entrance in Germany," in *Youth Employment and Joblessness in Advanced Countries*, ed. by D. G. Blanchflower, and R. B. Freeman, chap. 10, pp. 381–425. The University of Chicago Press.

GROOT, W., AND H. MAASSEN VAN DEN BRINK (2000): "Overeducation in the labor market: a meta-analysis," *Economics of Education Review*, 19(2), 149–158.

HECKMAN, J. J. (1997): "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32, 441–462.

HECKMAN, J. J., H. ICHIMURA, AND P. TODD (1997): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261 – 294.

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467 – 476.

MAIER, M., F. PFEIFFER, AND W. POHLMEIER (2003): "Skill Obselescence and Individual Heterogeneity," in this volume.

ROBERT, C. P., AND G. CASELLA (1999): *Monte Carlo Statistcal Methods.* Springer-Verlag, New York.

ROSENBAUM, P. R., AND D. B. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41 – 55.

ROY, A. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135 – 146.

RUBIN, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies," *Journal of Educational Psychology*, 66, 688 – 701.

SOEP GROUP (2001): "The German Socio-Economic Panel (GSOEP) After More Than 15 Years - Overview," *Vierteljahreshefte zur Wirtschaftsforschung*, 70, 7 – 14.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables with Weak Instruments," *Econometrica*, 65, 557 – 586.

TANNER, M., AND W. WONG (1987): "The Calculation on the Posterior Distributions by Data Augmentation (with discussion)," *Journal of American Statistical Association*, 82, 528 – 550.

VELLING, J., AND F. PFEIFFER (1997): *Arbeitslosigkeit, inadäquate Beschäftigung, Berufswechsel und Erwerbsbeteiligung.* ZEW Dokumentation Nr. 97-02, Mannheim.

VYTLACIL, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341.

Appendix

Prior Specification

Defining $\beta = (\beta_0', \beta_1', \beta_D')' : k \times 1$, $\alpha = (\alpha_0, \alpha_1)' : 2 \times 1$ and $\sigma = (\sigma_0^2, \sigma_1^2)' : 2 \times 1$, our prior distribution consists of five independent components:

- $\alpha_0 \sim N\left(\bar{a}_0, \bar{A}_0\right),$

- $\alpha_1 \sim N\left(\bar{a}_1, \bar{A}_1\right),$

- $\beta \sim N\left(\bar{b}, \bar{B}\right),$

- $\sigma_0^2 \propto N\left(\bar{g}_0, \bar{G}_0\right) \times \mathbb{1}\left(\sigma_0^2 > 0\right),$

- $\sigma_1^2 \propto N\left(\bar{g}_1, \bar{G}_1\right) \times \mathbb{1}\left(\sigma_1^2 > 0\right),$

where $\mathbb{1}(X \in A)$ is the indicator function which is equal to 1 if X is contained in the set A.

In order to express prior ignorance we choose

$$\bar{a}_0 = \bar{a}_1 = \bar{b} = \bar{g}_0 = \bar{g}_1 = 0,$$

and

$$\bar{A}_0 = \bar{A}_1 = \bar{G}_0 = \bar{G}_1 = 10^4, \ B_0 = 10^4 \times I_k.$$

Sampling Algorithm

The posterior density is proportional to the product of the prior density and the likelihood function. The conditional distributions of the parameters and the latent variables can easily be derived. Specifically, we follow Chib and Hamilton (2000) and use a reduced blocking step in order to improve the efficiency of the algorithm.

1. Sample $(\alpha_0, \alpha_1, \sigma_0^2, \sigma_1^2)$ by first sampling (α_0, σ_0^2) from $\alpha_0, \sigma_0^2 | Y, D, \beta$ which is proportional to

$$f_N(\alpha_0 | \bar{a}_0, \bar{A}_0) f_N(\sigma_0^2 | \bar{g}_0, \bar{G}_0) \mathbb{1}\left(\sigma_0^2 > 0\right)$$

$$\times \prod_{D_i=0} \left\{ \frac{1}{\sqrt{\alpha_0^2 + \sigma_0^2}} \exp\left[-\frac{1}{2} \frac{(Y_{i0} - X_i \beta_0)^2}{\alpha_0^2 + \sigma_0^2}\right]\right.$$

$$\left. \times \Phi\left[\frac{-W_i \beta_D - \frac{\alpha_0}{\alpha_0^2 + \sigma_0^2}(Y_{i0} - X_i \beta_0)}{\sqrt{2 - \frac{\alpha_0^2}{\alpha_0^2 + \sigma_0^2}}}\right]\right\}.$$

Then sample (α_1, σ_1^2) from $\alpha_1, \sigma_1^2|Y, D, \beta$ which is proportional to

$$f_N(\alpha_1|\bar{a}_1, \bar{A}_1)f_N(\sigma_1^2|\bar{g}_1, \bar{G}_1)\mathbb{1}(\sigma_1^2 > 0)$$

$$\times \prod_{D_i=1} \left\{ \frac{1}{\sqrt{\alpha_1^2 + \sigma_1^2}} \exp\left[-\frac{1}{2}\frac{(Y_{i1} - X_i\beta_1)^2}{\alpha_1^2 + \sigma_1^2} \right] \right.$$

$$\left. \times \Phi\left[\frac{W_i\beta_D + \frac{\alpha_1}{\alpha_1^2+\sigma_1^2}(Y_{i1} - X_i\beta_1)}{\sqrt{2 - \frac{\alpha_1^2}{\alpha_1^2+\sigma_1^2}}} \right] \right\}.$$

2. If $D_i = 0$ sample the D_i^* independently for $i = 1, \ldots, n$ from a $N(W_i\beta_D, 2)$ distribution truncated at the right by 0, if $D_i = 1$ sample D_i^* from a $N(W_i\beta_D, 2)$ distribution truncated at the left by 0.

3. In the case of $D_i = 1$ sample the Y_{i0}^* independently from a $N(\mu_Y, \Sigma_Y)$, where

$$\mu_Y = X\beta_0 + \frac{\alpha_0\alpha_1(Y_{i1} - X_i\beta_1) + \alpha_0\sigma_1^2(D_i^* - W_i\beta_D)}{\alpha_1^2 + 2\sigma_1^2},$$

$$\Sigma_Y = \sigma_0^2 + \frac{\alpha_0\sigma_1^2}{\alpha_1^2 + 2\sigma_1^2}.$$

If $D_i = 0$ sample the Y_{i1}^* independently from a $N(\mu_Y, \Sigma_Y)$ with

$$\mu_Y = X_i\beta_1 + \frac{\alpha_0\alpha_1(Y_{i0} - X_i\beta_0) + \alpha_1\sigma_0^2(D_i^* - W_i\beta_D)}{\alpha_0^2 + 2\sigma_0^2},$$

$$\Sigma_Y = \sigma_1^2 + \frac{\alpha_1\sigma_0^2}{\alpha_0^2 + 2\sigma_0^2}.$$

4. Sample β from its conditional distribution which is $\beta \sim N\left(\hat{\beta}, \hat{H}\right)$, where

$$\hat{H} = \left(\bar{B}^{-1} + \sum_{i=1}^{n} X_i'\Sigma^{-1}X_i \right)^{-1},$$

$$\hat{\beta} = \hat{H}\left(\bar{b}\bar{B}^{-1} + \sum_{i=1}^{n} X_i'\Sigma^{-1}Z_i \right),$$

and

$$\Sigma = \begin{pmatrix} \alpha_0^2 + \sigma_0^2 & \alpha_0\alpha_1 & \alpha_0 \\ \alpha_0\alpha_1 & \alpha_1^2 + \sigma_1^2 & \alpha_1 \\ \alpha_0 & \alpha_1 & 2 \end{pmatrix}.$$