College Education and Wages in the U.K.: Estimating Conditional Average Structural Functions in Nonadditive Models with Binary Endogenous Variables*

Tobias J. Klein†

first version: March 12, 2004
draft: October 27, 2006

Abstract

We propose and implement an estimator for identifiable features of correlated random coefficient models with binary endogenous variables and nonadditive errors in the outcome equation. It is suitable, e.g., for estimation of the average returns to college education when they are heterogeneous across individuals and correlated with the schooling choice. The estimated features are of central interest to economists and are directly linked to the marginal and average treatment effect in policy evaluation. They are identified under assumptions weaker than typical exclusion restrictions used in the context of classical instrumental variables analysis. In our application for the U.K., we relate levels of expected wages to unobserved ability, measured ability, family background, type of secondary school, and the decision whether to attend college.

JEL Classification: C14, C31, J31.

Keywords: returns to college education, correlated random coefficient model, local instrumental variables, local linear regression.

*I am especially grateful to Erich Battistin, Richard Blundell, Pedro Carneiro, Andrew Chesher, Johannes Gernandt, Stefan Hoderlein, Enno Mammen, Melanie Schienle, and Edward Vytlacil for stimulating discussions and helpful comments on earlier versions of this paper. Moreover, I would like to thank Richard Blundell, Lorraine Dearden, as well as Leslie McGranahan for sharing their data. Finally, I would like to thank UCL for its hospitality during the academic year 2003/4 and the European Commission for its financial support through the Marie Curie program. The paper was presented at the 2005 IAB conference "The Empirical Evaluation of Labour Market Programmes" in Nuremberg, the 2006 North American Winter Meeting of the Econometric Society in Boston and the 2006 ZEW Evaluation conference in Mannheim and I would like to thank the audiences for valuable comments. Finally, I would like to express my gratitude to two anonymous referees of the Junior Economits Peer-reviewed Discussion Paper Series (JEPS) for their detailed reports.

†University of Mannheim. *Address:* University of Mannheim, Department of Economics, Chair Prof. Stahl, Ph.D., 68131 Mannheim, Germany. *E-Mail:* klein@econ.uni-mannheim.de.

1. Introduction

In many econometric applications, the characterization of the impact of binary variables on an outcome variable is of central interest. Examples are the impact of an additional year of schooling, or attending college, on wages, or the impact of participation in a labor market program on unemployment duration. Often, conditional on observable covariates, these effects are considered to be heterogeneous across individuals and possibly correlated with the binary choice variable. This is true if the choice is based upon knowledge of the outcome that is superior to what is observed in the data.

In this paper, we model the log of individual wages, which we denote by Y, by a correlated random coefficient model of the form

$$Y = X'\varphi(D, U, V) \tag{1}$$

$$D = \mathbb{1}\{P(Z) \geq V\}, \tag{2}$$

where D is the binary endogenous variable. It is equal to one if the individual graduated from college. X is a K-vector of observable covariates in the wage equation, (1). In our application, we exploit a uniquely rich birth cohort data set, the British National Child Development Survey (NCDS), and include in X, among other variables, the type of secondary school that was attended, the social class of the parents, as well as other family background variables and accurately measured ability test scores at the age of 7 and 11. Z is a vector of covariates in the selection equation, (2), and includes the variables in X as well as the father's interest in the education of the child for which we assume that it can be excluded from the wage equation. As we will see below, such an exclusion restriction is not necessary for identification in our model but yields additional identifying power. X and Z include a constant as their respective first elements. U is a vector and represents "luck", and V is a scalar entering both the wage and the selection equation. It represents unobservable costs, benefits, and most importantly, talent and unobserved ability which we suppose to have an impact on both wages and the decision whether to attend college. Modelling this link is of economic interest and important once we aim at estimating the impact of changes in D or X on Y from non-experimental field data. As for stochastic restrictions, we assume that (U, V) are jointly independent of (X, Z) and that U is independent of V. This implies that $P(Z)$ is identified from observations.

In this model, *ceteris paribus* effects of changes in observables, D and X, on wages depend on unobservables U and V. In general, the model is not identified.[1] However, we will show that

[1] The model in (1) and (2) is nonadditive in the unobservables. Moreover, the vector X, in principle, could include approximating functions in a way such that the number of approximating functions grows with the sample size. Then, along with Newey (1997), (1) could be interpreted as a series approximation of a general nonseparable structural equation $Y = g(X, D, U, V)$. Together with (2) this is a triangular structure similar to the ones considered by Chesher (2003) and Imbens and Newey (2003). The key difference, however, is that here (2) is not invertible in V and hence identification fails since V enters as an argument. Chesher (2005) shows that in this case set identification may still be feasible.

under suitable conditions the expected level of wages, Y, for a given D, X, and V,[2]

$$\mathbb{E}[Y|D = d, X = x, V = v] = x'\mathbb{E}[\varphi(d, U, v)]$$

is identified from observations. We will refer to this identifiable feature of the wage equation as the *conditional average structural function* (CASF) and to $\mathbb{E}[\varphi(d, U, v)]$ as the vector of conditional average *ceteris paribus* effects, understanding the notion of *ceteris paribus* as holding all other factors constant, including V, averaging only over U. We believe that not only average *ceteris paribus* effects, where we average over V and U, are of interest but also their dependence on V. In Section 3, we define the parameters of interest more formally and link them to a variety of treatment effect parameters that are considered in the literature on program evaluation.

Heckman and Vytlacil (1999, 2000a, 2000b; HV in the remainder) establish nonparametric identification of the CASF under weaker stochastic restrictions. They assume that (U, V) are jointly independent of Z conditional on X instead of our assumption that (U, V) are jointly independent of (X, Z). However, they require conditions on the support of $P(Z)$ to hold conditional on X, whereas we require them to hold only unconditionally.

To illustrate this point, nonparametric estimation of the CASF as suggested by the identification result of HV is not feasible in our application because not only the expected level of wages needs to be estimated conditional on $X = x$, $D = d$, and $P(Z) = p$, but also its partial derivative with respect to p which requires continuous variation of $P(Z)$ conditional on D and X, a requirement that is not met in our data because there is no such continuous variable in Z. This shows that there is a tradeoff between flexibility of the model—the model by HV is fully nonparametric—and data requirements.

Carneiro, Heckman, and Vytlacil (2005) and Carneiro and Lee (2005) propose what we shall refer to as the additive model. Write X as $(1, X'_{-1})'$. Then, instead of (1), which can be written as

$$Y = \varphi_1(D, U, V) + X'_{-1}\varphi_{-1}(D, U, V),$$

they consider a wage equation of the form

$$Y = \mu(D, U, V) + X'_{-1}\gamma(D, U)$$

and show that the CASF is identified under the same stochastic restrictions and support conditions that we use in this paper.

One limitation of their model is that they do not allow for the effect of X on Y to depend on V. This nonseparability is an important aspect of unobserved heterogeneity and is of economic interest in many applications with binary endogenous variables. Estimates for the more general model that is proposed in this paper indicate that these nonseparabilities are present in our data. For example, the expected effect of the parents' social class measured by the father's occupation depends on the level of unobserved ability, V. Moreover, we find that imposing separability results in biases which are significant.

We estimate the model by local linear smoothing. Our estimator is a *local instrumental*

[2]We will denote (vectors of) random variables by uppercase letters and their respective typical elements by lowercase letters.

variables estimator built on the conventional two stage least squares instrumental variables (IV) estimator, except that we let the coefficients depend on the value of $P(Z)$. It turns out that it is substantially easier to implement than the estimator used for the additive model.[3]

Our results indicate that returns of attending college relative to obtaining just A-levels to be sizable. Moreover, we find evidence for heterogeneity of monetary returns. They are lower for individuals who actually attend college as compared to the returns for those who don't. This can be traced back to both observable and unobservable factors, and the interaction of the two. One finding is that returns are increasing in the father's years of education. Unlike other studies, we don't find clear cut evidence for sorting based on comparative or absolute advantage.

The paper is organized as follows. In Section 2, we embed our study into the literature. The full characterization of the econometric model, the identification result, and the proposed estimator are presented in Section 3. Section 4 contains the results from the empirical analysis. Section 5 concludes.

2. Related Results

In this section, we briefly relate our model to the literature.[4] Furthermore, we discuss aspects of modelling unobserved heterogeneity, and most importantly ability.

In our application, we model two types of ability. The first consists of math and verbal ability test scores at the age of 7 and 11, which we include in our set of covariates. The availability of this information is a key advantage of the NCDS since in many other data sets, e.g. the Family Expenditure Survey, the General Household Survey, or the Labor Force Survey, such precisely measured information is not available. Blundell, Dearden, and Sianesi (2005, BDS in the remainder) analyze the same data using OLS, IV, matching, and control function techniques. In Section 4, we compare our estimates of average returns to the ones of BDS.

The second type of ability is contained in V which enters both the wage and the selection equation. In the statistics literature, V is sometimes referred to as a confounding variable, see e.g. Fisher (1935, Ch. 7) and Yates (1937). For simplicity, we refert to V as unobservable ability.

This is well in line with the economics literature, where the term "ability" is often used in different contexts and with different meanings. Griliches (1977, p. 7) defines it as "an unobserved latent variable that both drives people to get relatively more schooling and earn more income, given schooling, and perhaps also enables and motivates people to score better on various tests." Along those lines, Taubman and Wales (1972) and Taubman (1973) call it "mental ability" and Willis and Rosen (1979) use the expression "talent". On the other hand, Griliches (1977, p. 8) suggests that one could also interpret ability as "initial human capital".

[3]First applications are Carneiro, Heckman, and Vytlacil (2005), Carneiro and Lee (2005) and Heckman, Urzua, and Vytlacil (2004) where the model is estimated using a double residual regression involving several additional steps, see Robinson (1988).

[4]The question of how to estimate the returns to schooling and college education, which is closely related to the estimation of respective counterfactual wage levels, is one of the classical questions in econometrics. For two excellent surveys of the literature on the returns to schooling see Griliches (1977) and Card (2001). For an early survey on the returns to college education see Solmon and Taubman (1973).

More broadly, Becker (1967) elaborates on whether there are several types of ability and Willis and Rosen (1979, p. S29) note that ability is potentially multi factoral.

In fact, in our model, V is the projection of *all* unobservable factors that are common to the wage and the selection equation onto a scalar. Ashenfelter and Mooney (1968), Griliches and Mason (1972), Hansen, Weisbrod, and Scanlon (1970), Weisbrod and Karpoff (1968) and Leibowitz (1974) discuss and partly analyze the link between ability as well as other factors and earnings in more detail. Examples of such other factors include wealth, parent's income, status, social origin, motivation, quality of schooling, and idiosyncratic preferences. Here, we reach a natural limitation of our data since not all of those factors are observable. We proxy for some of these factors by including accurately measured family background variables, that are contained in the NCDS, in our set of covariates so that only the remaining variation is captured by V if it is common to the wage and the selection equation, and U if its only impact is on wages.

In general, econometric challenges arise from the fact that, *via* what we call unobserved ability, V, the return to schooling and college education is likely to be correlated with schooling and college choice once it results from optimizing behavior by economic agents who act on their knowledge of their ability. This gives rise to the classical selection problem in econometrics which could be overcome relatively easily if a perfect measure of ability was available, for example by including this measure into the set of regressors in the wage equation. Griliches (1977) discusses econometric consequences when an imperfect measure is used, i.e. when ability is measured with error. Along these lines, Chamberlain (1977) argues that it is instructive to think of unobserved ability as being a left-out variable.

Early contributions discussing the selection problem in detail include Heckman (1978), Heckman and Robb (1985, 1986) and Willis and Rosen (1979). A variety of approaches to this challenging problem has been taken over the last four decades. Identifying assumptions include parametric assumptions, as well as conditional (mean) independence and monotonicity in order to identify mean returns. Also, quantile invariance has proved to be a powerful identifying assumption.[5]

Most of these approaches rely on the presence of IVs that can be excluded from the earnings equation. IVs that have been used are quarter of birth Angrist and Krueger (1991) and parental interest in education (BDS) as well as, e.g., the level of tuition fees, distance to college, and parental education, see Card (2001) for details. Angrist and Krueger (2001) advocate the use of natural experiments such as institutional changes as instruments giving rise to variation exogenous to the earnings equation. In our application, we derive *additional* identifying power from the assumption that the father's interest in the education of the child can be excluded from the wage equation.

[5] For distributional assumptions see, e.g., Heckman (1978), and **?**. Conditional independence is assumed in Rosenbaum and Rubin (1983). Heckman and Vytlacil (1998) exploit additivity of the error term in a random coefficient framework. Imbens and Angrist (1994), Angrist, Graddy, and Imbens (2000), HV, Carneiro, Heckman, and Vytlacil (2005), Heckman and Vytlacil (2005), and Abadie, Angrist, and Imbens (2002) exploit monotonicity, which is implied by the selection model. In Section 3, it will become clear that this is what we do in this paper as well. Quantile invariance is relied on in Chernozhukov, Imbens, and Newey (2004) and Chernozhukov and Hansen (2005). It is well beyond the scope of this paper to review the literature. However, the reader is referred to, e.g. Blundell and Powell (2003) for the relationship between IV and control function estimators, HV as well as Heckman and Vytlacil (2005) for the relationship between estimators based on monotonicity and classical IV estimators and OLS, and BDS for a comparison of OLS, IV, matching and control function estimators.

The approach we take in this paper has several key advantages. First, we do not restrict selection to be based solely on observables which underlies OLS regressions, classical IV estimation, the random coefficient model suggested by Heckman and Vytlacil (1998), and matching.[6] Second, we do not have to specify the joint distribution of unobservables which underlies parametric approaches. Third, our model is nonparametric in the dimension of the unobserved heterogeneity since the dependence of the random coefficients on D and V is not constrained by functional form assumptions. Forth, as we have already discussed in the introduction, data requirements are weaker than in the fully nonparametric setup of HV, and equal to the ones of the additive model of Carneiro, Heckman, and Vytlacil (2005). At the same time, our model is more general in the sense that we allow both the random coefficient and D to depend on V.

3. Econometric Approach

This section contains the formal results underlying our analysis in Section 4. Our point of departure is the correlated random coefficients model that was given in (1) and (2). We restate it for convenience:

$$Y = X'\varphi(D, U, V) \tag{1}$$

$$D = \mathbb{1}\{P(Z) \geq V\}. \tag{2}$$

(1) is the wage equation and (2) is the selection equation. We impose the following stochastic restrictions.

ASSUMPTION 1 (Stochastic Restrictions): (i) (U, V) *are jointly independent of* (X, Z) *and* (ii) U *is independent of* V.

This allows Z to contain variables also included in X and *vice versa*. Assumption 1(i) requires the unobservables (U, V) to be jointly independent of the observables (X, Z). This is considerably weaker than the IV type assumption that D is independent of the unobservables in the outcome equation conditional on Z and X. Assumption 1(ii) restricts the randomness in Y through U to be completely random so that U represents luck, whereas V can be thought of as a confounding factor.[7]

[6] These models assume that conditional on observables, D is independent of either the effect from changes in D, or the error term in the outcome equation, or both. Garen (1984), Heckman (1978), Newey, Powell, and Vella (1999) as well as Pinske (2000) and Blundell and Powell (2003) pursue a control function approach. Imbens and Newey (2003) generalize this approach. Newey and Powell (2003), Darolles, Florens, and Renault (2003), and Das (2005) investigate the case in which the error term is additive. Notice that in our case identification is complicated by the fact that the endogenous variable is binary so that a control function approach in which we include the first stage residual into the second stage is not feasible because the selection equation is not invertible in V. It will become clear in Section 3 that the estimation step in our approach boils down to estimation of the expected outcome conditional on D, X, and $P(Z)$. Identification of the parameters of interest is achieved by exploiting the monotonicity implied by the selection model.

[7] Assumption 1(ii) is not restrictive. $\varphi(D, U, V)$ is a nonparametric function of the observable D and unobservables (U, V). Therefore, it can at most be identified up to normalizations on the joint distribution of unobservables.

Apart from the stochastic restrictions we assume that the following regularity conditions hold.

ASSUMPTION 2 (Regularity Conditions): *(i) All first moments exist and (ii) the distribution of V is absolutely continuous with respect to Lebesgue measure.*

Assumption 2(i) ensures that all parameters of interest are well defined. Assumption 2(ii) implies that V is a continuous random variable. This allows us, w.l.o.g., to normalize V from now on to be uniformly distributed on the unit interval, see, e.g., Vytlacil (2002) for details. From Assumption 1(i) it follows immediately that $P(Z)$ is identified from observations since it is equal to $\Pr(D = 1|Z)$. For simplicity, we will write P for $P(Z)$ in the remainder, with typical element p.

3.1. Parameters of Interest

We have already argued in the introduction that the CASF,

$$\mathbb{E}[Y|D = d, X = x, V = v] = x'\mathbb{E}[\varphi(d, U, v)]$$

is of special interest in our application. The terminology we use was introduced by Blundell and Powell (2003) who suggest to focus on the average structural function, $\mathbb{E}[Y|D = d, X = x]$. Likewise, Imbens and Newey (2003) call it the average conditional response. Following Goldberger (1972), who calls an equation *structural* if it represents a causal link rather than a mere empirical association, we prefer to think of the wage equation as a structural equation. We believe that for a given D and X the *dependence* of the average structural function on V is of central economic interest by itself and hence focus on the average *conditional* on V.

A second object of interest that is related to the CASF is the conditional average *ceteris paribus* effect of changes in X_k, e.g. the type of secondary school that was attended or the social class of the father, for a given D, X_{-k}, and V,[8]

$$\frac{\partial \mathbb{E}[Y|D = d, X = x, V = v]}{\partial x_k} = \mathbb{E}[\varphi_k(d, U, v)].$$

Moreover, we are interested in the expected *ceteris paribus* effect of changes in D for a given X

Assume that the joint distribution of unobservables is absolutely continuous with respect to Lebesgue measure. Then, the restrictions on the joint distribution of observables imposed by any joint distribution of $(\widetilde{U}, \widetilde{V})$ are the same as the ones imposed by the joint distribution of (U, V), where $v = F_{\widetilde{V}|\widetilde{U}}(\widetilde{v})$ with V being uniformly distributed independently of U. For example, we could have $U = \widetilde{U}$ or any positive monotone transformation thereof. See also Imbens and Newey (2003) for a related discussion.

[8]The kth element of a vector x is denoted by x_k. The remaining elements are denoted by x_{-k}. For discrete X_k the partial derivative is replaced by an appropriate difference.

and V,

$$\mathbb{E}[Y|D = 1, X = x, V = v] - \mathbb{E}[Y|D = 0, X = x, V = v]$$
$$= x'(\mathbb{E}[\varphi(1, U, v)] - \mathbb{E}[\varphi(0, U, v)]).$$

This is Björklund and Moffit's (1987) marginal treatment effect. It is the expected effect of a college degree on wages for a given level of unobserved ability and for a given vector of covariates. The well-known average treatment effect, averaged over the population distribution of unobserved ability, for a given $X = x$ is given by

$$x' \int_0^1 (\mathbb{E}[\varphi(1, U, v)] - \mathbb{E}[\varphi(0, U, v)])dv, \tag{3}$$

recalling that we have normalized V to be uniformly distributed.

3.2. Identification

In this subsection, we show identification of the CASF at a given D, X, and V under Assumption 1 and 2. The estimator we implement, which is built on local linear smoothing, is proposed thereafter.

Because of the multiplicative structure of the wage equation, identification of the CASF at $D = d$, $V = v$, and any $X = x$ is equivalent to identification of the conditional average *ceteris paribus* effects. The average structural function as well as average *ceteris paribus* effects are identified at $D = d$ if the CASF is identified at all V in the open unit interval, recalling that we have normalized its distribution to be uniformly distributed and that the endpoints have probability measure zero. Finally, if the (conditional) average structural function is identified at both $D = 0$ and $D = 1$, the average (marginal) treatment effect is as well.

From the model in (1), it follows that

$$\mathbb{E}[Y|D = 1, P = p, X = x] = x'\mathbb{E}[\varphi(1, U, V)|D = 1, P = p, X = x] \tag{4}$$

which is equal to

$$x'\mathbb{E}[\varphi(1, U, V)|P \geq V, P = p, X = x]$$

by the selection model in (2). But this is

$$x'\mathbb{E}[\varphi(1, U, V)|X = x, p \geq V].$$

By Assumption 1(i) we get that this is equal to

$$\mathbb{E}[x'\varphi(1, U, V)|p \geq V] = x'\mathbb{E}[\varphi(1, U, V)|p \geq V] =: x'\beta(1, p).$$

Note that $\mathbb{E}[\varphi(1, U, V)|p \geq V]$ is a function of p which we will denote by $\beta(1, p)$ in the remainder. Since the left hand side of (4) is identified from observations at points of support $X = x$ and $P = p$, $\beta(1, p)$ is identified if we observe at least K linearly independent values of X for $D = 1$ (rank condition). $\beta(0, p)$ is defined accordingly and a similar result holds for $D = 0$.

Starting from this, we show that the CASF is identified. We state the result in a theorem which resembles Lemma 1 from Carneiro and Lee (2005). Following HV, they show nonparametric identification under weaker stochastic restrictions than the ones in Assumption 1, at the price of stronger support conditions that need to hold for their result. Only when they estimate the model they impose the restrictions in Assumption 1. We show the proof for two reasons. First, strictly speaking, our identification result is not implied by their Lemma 1, even though their proof is similar to ours. Second, our rank condition differs from theirs.

We call p a limit point of the support of P, if P has a continuous density in a neighborhood around p which is bounded away from zero. Note that at $P = p$ derivatives of differentiable functions of P are identified from observations.

Theorem 1 (Identification): *Assume that $\beta(0, p)$ and $\beta(1, p)$ are continuously differentiable with respect to p and that we observe at least K linearly independent realizations of X for every D and $P = p$ (rank condition). Then, under Assumptions 1 and 2 the CASF is identified at $V = p$, where p is a limit point of the support of P, and given by*

$$x'\mathbb{E}[\varphi(0, U, p)] = x'\left(\beta(0, p) - (1 - p)\frac{\partial\beta(0, p)}{\partial p}\right)$$
$$x'\mathbb{E}[\varphi(1, U, p)] = x'\left(\beta(1, p) + p\frac{\partial\beta(1, p)}{\partial p}\right).$$

Proof. We prove identification of $\mathbb{E}[\varphi(D, U, V)|D = 1, V = p]$. The proof for $\mathbb{E}[\varphi(D, U, V)|D = 0, V = p]$ is similar. Recall that we have normalized V to be uniformly distributed. By definition,

$$x'\mathbb{E}[\varphi(1, U, V)|p \geq V] = x'\beta(1, p).$$

From the normalization on V and Assumption 1(ii) it follows that

$$x' \int_0^p \int_{-\infty}^{\infty} \varphi(1, u, v)\, \mu(du)\, dv/p = x'\beta(1, p), \tag{5}$$

where $\mu(du)$ is the marginal probability measure of u. Multiplying both sides by p gives

$$x' \int_0^p \int_{-\infty}^{\infty} \varphi(1, u, v)\, \mu(du)\, dv = x'\beta(1, p)p$$

and differentiating both sides with respect to p using Leibnitz' rule reveals that

$$x' \int_{-\infty}^{\infty} \varphi(1, u, p)\, \mu(du) = x'\beta(1, p) + px'\frac{\partial\beta(1, p)}{\partial p}.$$

If p is a limit point of the support of P both $\beta(1, p)$ and $\partial\beta(1, p)/\partial p$ are identified from observations at $P = p$. The left hand side is the object of interest. □

Finally, notice that the proof relies on the monotonicity of D in P implied by the selection model which allows us to formulate (5). See also Klein (2006) for a discussion and an analysis of the case in which monotonicity does not hold, but is wrongly assumed.

3.3. Estimation

We have established in our discussion that from the model and the conditions of Theorem 1 it follows that

$$\mathbb{E}[Y|D = d, P = p, X = x] = x'\beta(d, p) \ , \ d \in \{0, 1\},$$

where $\beta(d, p)$ is a coefficient vector with coefficient functions $\beta_k(d, p)$, $k = 1, \ldots, K$. Both depend on the observable D, and P which is identified from observations. This is a version of the varying coefficient model which was suggested by Cleveland, Grosse, and Shyu (1991) and Hastie and Tibshirani (1993).

In a first step, we parametrically estimate P. For the second step we assume that the coefficient functions are bounded and have bounded second derivatives which allows us to estimate them by local linear smoothing. See, for example Fan and Zhang (1999) and Xia and Li (1999) for details as well as a proof of consistency and results on rates of convergence of the estimator. This estimation procedure is usually motivated by a Taylor expansion of the coefficient function in $\tilde{p}$ about $\tilde{p} = p$ which yields

$$\beta_k(d, \tilde{p}) = \beta_k(d, p) + \frac{\partial \beta_k(d, p)}{\partial p}(\tilde{p} - p) + \frac{1}{2}\frac{\partial^2 \beta_k(d, \bar{p})}{\partial p^2}(\tilde{p} - p)^2,$$

where $\bar{p}$ is a point between p and $\tilde{p}$. We select all observations with $D = d$ and index them by i, $i = 1, \ldots, n$. Our estimator of $\beta(d, p)$ and $\partial\beta(d, p)/\partial p$ is the solution of a and b to the following minimizer

$$\arg\min_{a,b} \left\{ \sum_{i=1}^{n} K\left(\frac{p_i - p}{h}\right) \cdot \left(y_i - \begin{bmatrix} x_i \\ (p_i - p) \cdot x_i \end{bmatrix}' \begin{pmatrix} a \\ b \end{pmatrix}\right)^2 \right\},$$

where $K(\cdot)$ is a kernel function with the usual properties and h is the bandwidth. Since fitted values p_i were parametrically estimated in a first step we do not expect them to have an impact on the distribution of the second step estimator in a first order asymptotic sense. However, we obtain confidence intervals, accounting for the first step estimation error, using a bootstrap procedure.

Estimates of the objects of interest can be obtained from these estimates of $\beta(d, p)$ and $\partial\beta(d, p)/\partial p$ using the formulas from Theorem 1.

4. College Education and Wages in the U.K.

4.1. Data

We implement the estimation procedure which was proposed in Section 3 for U.K. data from the NCDS. The NCDS is conducted by the Centre for Longitudinal Studies at the Institute of Education in London. It is a longitudinal data set and keeps detailed records for all those living in Great Britain who were born between March 3 and 9, 1958. The data were collected in 1958, in 1965 (when members were aged 7 years), in 1969 (age 11), in 1974 (age 16), in 1981 (age

23), in 1991 (age 33) and 1999-2000 (age 41-42). The NCDS has gathered data from respondents on child development from birth to early adolescence, child care, medical care, health, physical statistics, school readiness, home environment, educational progress, parental involvement, cognitive and social growth, family relationships, economic activity, income, training, and housing.

Recently, BDS study these data using IV estimation, a control function estimator, and matching techniques. For a more detailed data description and variable definitions the reader is referred to their paper.

Their, as well as our, outcome of interest is log hourly wages in 1991, this is at the age of 33. We select individuals who at least completed their A levels, from which 51.4% are higher education graduates. We say that an individual completes his A levels if he completed at least one A level which is generally obtained at the end of secondary school, see BDS for details. Notably, we distinguish between college graduates ($D = 1$), who have completed some kind of higher education, and those who have obtained A levels only ($D = 0$). We focus on employed males and select individuals with non-missing verbal and math ability test scores. This leaves us with 1501 observations.

The NCDS contains a host of accurately measured variables including information about the type of secondary school that was attended and a number of family background variables. In the U.K., secondary school is attended from the age of 11 to 12 on for 7 years. The individuals in our sample were born in 1958 so that they entered secondary school in the late 1960s. At that time the public school system was changing. Until then, there were two basic types of public secondary schools in the U.K., Secondary modern and Grammar schools. Secondary modern schools were intended for children who would be going into a trade and focussed on practical skills. Grammar schools were intended to prepare pupils for higher education. In the 1960s, comprehensive schools were promoted as an alternative and started to partly replace the old system providing complete and general education. But in fact, which route was pursued for the school system highly depended on the respective local authority. Nowadays, there is a mixture of types of public schools. Alongside public schools there are prestigious Private schools such as Eton college, which are sometimes still referred to as “public schools” since they are open for the paying public as opposed to a religious school.[9]

In our analysis we proxy social class by the type of occupation of the father when the child was 16. Categories are professional, intermediate, skilled and semi-skilled non-manual as well as skilled or semi-skilled manual, and unskilled.

Table 1 contains summary statistics for our data. Notably, wages are higher for college graduates, and as compared to college non-graduates more college graduates (i) went to Grammar or Private school, (ii) have a father who is professional or intermediate, and (iii) have better educated parents on average.

4.2. First Stage Estimates

The first stage of our two stage estimator consists of fitting values of P by estimating a probit model. Our set of variables in the selection equation, Z, consists of the parent’s interest in the

[9]See, e.g., http://en.wikipedia.org/wiki/Education_in_the_United_Kingdom (February 2006).

	no college (49%)		college (51%)	
	mean	std.	mean	std.
LOG HOURLY WAGE AT THE AGE OF 33	2.04	0.40	2.32	0.37
FATHER'S INTEREST IN THE EDUCATION OF THE CHILD AT THE AGE OF 7				
expects too much	0.01	0.07	0.03	0.17
very interested	0.28	0.45	0.43	0.50
some interest	0.24	0.43	0.23	0.42
MOTHER'S INTEREST IN THE EDUCATION OF THE CHILD AT THE AGE OF 7				
expects too much	0.03	0.17	0.05	0.21
very interested	0.38	0.49	0.56	0.50
some interest	0.44	0.50	0.31	0.46
ABILITY MEASURES				
math ability at 7	55.17	23.97	66.40	21.53
math ability at 11	57.39	16.06	67.25	14.26
verbal ability at 7	80.15	20.74	90.45	13.32
verbal ability at 11	59.52	20.52	72.05	16.98
SECONDARY SCHOOL TYPE				
Secondary Modern	0.15	0.36	0.09	0.29
Comprehensive school	0.52	0.50	0.42	0.49
Grammar	0.08	0.28	0.21	0.41
Private	0.04	0.20	0.10	0.29
other	0.02	0.14	0.01	0.11
missing school information	0.19	0.39	0.17	0.37
SOCIAL CLASS OF THE FATHER				
professional	0.03	0.17	0.10	0.30
intermediate	0.16	0.37	0.23	0.42
skilled non-manual	0.09	0.29	0.09	0.29
skilled manual	0.34	0.48	0.25	0.44
semi-skilled non-manual	0.01	0.09	0.01	0.07
semi-skilled manual	0.09	0.29	0.06	0.23
unskilled	0.19	0.39	0.18	0.39
missing/unemployed/no father	0.08	0.27	0.08	0.28
FAMILY BACKGROUND VARIABLES WHEN THE CHILD WAS 16				
father's years of education	7.84	4.32	8.35	5.11
missing	0.21	0.41	0.23	0.42
mother's years of education	8.04	4.17	8.20	4.65
missing	0.20	0.40	0.22	0.41
father's age	44.17	11.69	45.06	10.89
missing	0.05	0.22	0.04	0.19
mother's age	42.41	8.76	42.75	8.83
missing	0.02	0.15	0.03	0.16
mother was employed	0.58	0.49	0.55	0.50
number of siblings	1.70	1.57	1.47	1.41
REGION WHEN THE CHILD WAS 16				
North Western	0.10	0.30	0.10	0.30
North	0.07	0.26	0.08	0.27
East and West Riding	0.06	0.24	0.07	0.25
North Midlands	0.08	0.28	0.07	0.25
Eastern	0.07	0.26	0.09	0.28
London and South East	0.14	0.34	0.14	0.35
Southern	0.06	0.23	0.07	0.25
South Western	0.07	0.26	0.07	0.25
Midlands	0.09	0.28	0.07	0.26
Wales	0.06	0.23	0.06	0.23
other	0.20	0.40	0.19	0.39

Table 1: Summary Statistics.

education of the child, math and reading ability test scores at the age of 7 and 11, indicator variables for secondary school type, the father's social class when the child was 16, as well as other family background variables and, in some specifications, region. As was shown in Section 3 an exclusion restriction is not needed for identification of the CASF in our model—unlike for nonparametric identification as in HV.

Note that whereas the interpretation of the estimated probit coefficients as *ceteris paribus* effects heavily relies on the distributional assumptions in a probit model, the fitted values of the propensity score are less sensitive to violations of those assumptions once we interpret the usual probit model as a reduced form.[10] As suggested by the literature, and in order to undertake a sensitivity analysis, we estimated these fitted values by ordinary least squares, see Kelejian (1971) and the discussion in Angrist and Krueger (2001). However, our results did not change qualitatively.

Table 2 contains coefficient estimates for 5 different specifications. Throughout, the direction of the impact is as expected and in line with the literature which takes a closer look at the channels though which parents' education is transmitted to the children, see Goldberger (1989) and Haveman and Wolfe (1995) for an overview and discussions. Column (1) is the full specification in which indicator variables for region were included. Column (2) is the same specification except that secondary school type was left out because it could arguably be endogenous. This is the case whenever conditional on measured ability and all other controls in Z, those who know already that they will be more likely to go to college attain a special kind of secondary school, e.g. Grammar school. The remaining coefficients are largely unchanged. In the first two specifications, the region indicator variables were all insignificant. Column (3) and (4) contain estimates obtained from the specification in (1) and (2), respectively, except that these indicator variables were left out. Again, in comparison to the first two columns, the estimates remained largely unchanged. For our final specification in column (5) we left out the mother's interest in the education of the child since it is highly correlated with father's interest.

Moreover, we left out some of the insignificant indicator variables for secondary school type, social class and family background. Our estimates show throughout that parents' interest has a significant impact on the probability of attending college, so do the ability measures, whether the child went to Grammar school, and whether the father is professional.

Figure 1 shows the the sample distributions of the fitted values of P. For both $D = 0$ and $D = 1$ the support is almost equal to the full unit interval. Note that the distributions differ between $D = 0$ and $D = 1$. This shows that the variables in Z have explanatory power.

4.3. Second Stage Implementation

In the second stage, drawing on Section 3's results we estimate the mean coefficient functions, $\beta(d, p)$ and their derivatives with respect to p. For smoothing in the direction of p we use an Epanechnikov kernel and estimated the coefficient vectors at 101 grid points between 0 and 1. As we have seen, this is a one-dimensional nonparametric problem. The bandwidths were chosen using a standard leave-one-out cross validation procedure. It turns out that the optimal bandwidth for $D = 0$ is infinitely large, implying estimation of a fully interacted model without

[10] Willis and Rosen (1979) invoke a set of assumptions which allows them to estimate both a reduced form and a structural probit.

	(1)		(2)		(3)		(4)		(5)	
	coeff.	*t*-stat.	coeff.	*t*-stat.	coeff.	*t*-stat.	coeff.	*t*-stat.	coeff.	*t*-stat.
FATHER'S INTEREST IN THE EDUCATION OF THE CHILD										
expects too much	1.12	3.09	1.11	3.11	1.13	3.15	1.13	3.18	1.23	3.72
very interested	0.13	1.16	0.12	1.10	0.12	1.11	0.11	1.04	0.27	3.17
some interest	0.26	2.73	0.25	2.61	0.26	2.68	0.24	2.57	0.22	2.47
MOTHER'S INTEREST IN THE EDUCATION OF THE CHILD										
expects too much	0.16	0.70	0.15	0.65	0.17	0.74	0.15	0.69		
very interested	0.20	1.44	0.21	1.57	0.22	1.59	0.23	1.70		
some interest	-0.02	-0.13	-0.01	-0.05	0.00	-0.01	0.01	0.05		
ABILITY MEASURES										
math ability at 7	0.00	1.90	0.00	2.17	0.00	1.98	0.00	2.28	0.00	2.24
math ability at 11	0.01	3.83	0.01	3.87	0.01	3.86	0.01	3.93	0.01	3.79
verbal ability at 7	0.01	3.97	0.01	4.16	0.01	4.01	0.01	4.15	0.01	4.04
verbal ability at 11	0.00	1.36	0.00	1.63	0.00	1.28	0.00	1.55	0.00	1.19
SECONDARY SCHOOL TYPE RELATIVE TO COMPREHENSIVE SCHOOL										
Secondary Modern	0.01	0.09			0.00	-0.02				
Grammar	0.27	2.32			0.27	2.36			0.29	2.71
Private	0.11	0.71			0.11	0.72			0.12	0.78
other	-0.33	-1.17			-0.32	-1.15				
missing school information	-0.07	-0.59			-0.07	-0.59				
SOCIAL CLASS OF THE FATHER RELATIVE TO UNSKILLED										
professional	0.57	2.63	0.62	2.84	0.58	2.65	0.62	2.86	0.45	2.56
intermediate	0.17	1.06	0.22	1.37	0.18	1.09	0.23	1.41	0.04	0.45
skilled non-manual	0.13	0.72	0.14	0.80	0.14	0.75	0.15	0.83		
skilled manual	0.16	1.05	0.17	1.09	0.16	1.06	0.17	1.10		
semi-skilled non-manual	-0.01	-0.03	0.01	0.02	0.03	0.07	0.06	0.13		
semi-skilled manual	0.14	0.75	0.14	0.72	0.15	0.78	0.14	0.76		
missing/unemployed/no father	0.14	0.63	0.04	0.23	0.07	0.37	-0.05	-0.31		
FAMILY BACKGROUND VARIABLES WHEN THE CHILD WAS 16										
father's years of education	0.02	0.39	0.02	0.37	0.02	0.38	0.02	0.37	0.03	1.42
missing	0.79	2.12	0.83	2.26	0.78	2.11	0.83	2.26	0.36	1.45
mother's years of education	0.03	0.48	0.03	0.53	0.02	0.41	0.02	0.45		
missing	-0.47	-1.21	-0.43	-1.11	-0.48	-1.25	-0.44	-1.15		
father's age	0.00	-0.10	0.00	-0.11	0.00	-0.21	0.00	-0.22		
missing	-0.49	-0.95	-0.48	-0.94	-0.55	-1.07	-0.54	-1.06		
mother's age	0.02	1.19	0.01	1.11	0.02	1.23	0.01	1.14		
missing	0.97	1.68	0.93	1.62	0.99	1.74	0.96	1.68		
mother was employed	-0.09	-1.04	-0.09	-1.08	-0.09	-1.05	-0.09	-1.06	-0.08	-1.00
number of siblings	-0.04	-1.58	-0.04	-1.56	-0.05	-1.60	-0.04	-1.59	-0.04	-1.48
interaction father's education x age	0.00	0.65	0.00	0.78	0.00	0.64	0.00	0.75		
interaction mother's education x age	0.00	-0.98	0.00	-0.99	0.00	-0.94	0.00	-0.93		
REGION WHEN THE CHILD WAS 16										
indicator variables	yes		yes		no		no		no	
CONSTANT	-3.34	-5.69	-3.48	-6.03	-3.21	-5.62	-3.32	-5.94	-2.59	-8.40
McFadden *R*-squared	0.14		0.14		0.14		0.14		0.13	

Table 2: First stage probit coefficient estimates.

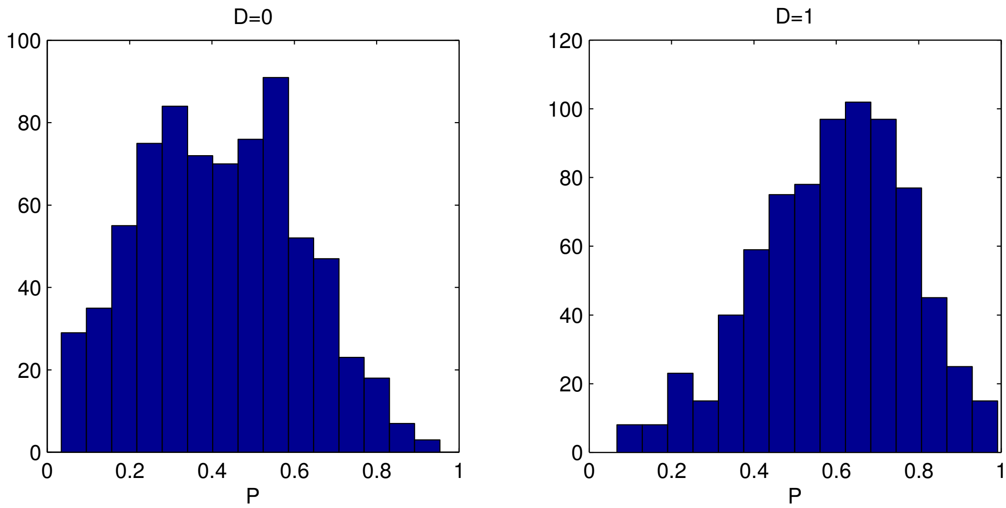

Figure 1: Sample distribution of the propensity score conditional on D.

any smoothing. For $D = 1$ the optimal bandwidth is 1.7. The required rank conditions hold in our data, i.e. the weighted $n \times 2K$ matrix of explanatory variables and interaction terms is of rank $2K$ at all evaluation points p.

From these estimates, which we provide with hats in the remainder, we calculate the vector of conditional average *ceteris paribus* effects for a given d and v, $\widehat{\mathbb{E}}[\varphi(d, U, v)]$, and the CASF, $x'\widehat{\mathbb{E}}[\varphi(d, U, v)]$ as well as other identifiable features of interest. In our bootstrap procedure for respective confidence intervals we acknowledge the fact that the propensity score is estimated in a first step by estimating it within every one of 1,000 bootstrap replications. For illustration, Figure 2 in the appendix contains estimates of the CASF and the marginal treatment effect for a representative individual with median characteristics. In particular, this representative individual went to comprehensive school, its father has 9 years of education and is neither professional nor intermediate, his mother is employed, and he has 1 sibling. Next, we go though the results in detail.

4.4. Average Returns to College Education

We calculate average returns using (3), replacing x by respective population means. For the average treatment effect on the untreated and treated, we simulate the distribution of V conditional on D by exploiting the structure of the selection model. For example, if we observe an individual with $D = 0$ and $P = p$, we would draw values of V from a uniform distribution on $(p, 1]$. Respective confidence intervals account for the simulation error. In Table 3, we compare our estimates to estimates obtained from an OLS regression, two stage IV estimates, as well as matching estimates obtained by BDS. The OLS estimate can be interpreted as the average difference in earnings observed in the population once we control for differences in covariates. This observed difference in earnings can be traced back to a selection effect and a causal effect of a higher education degree. Not surprisingly, the OLS estimate is very close to the matching

	estimate	95% conf. int.	
ATE population	0.46	0.04	0.89
ATE treated	0.26	-0.11	0.64
ATE untreated	0.63	0.03	1.22
OLS	0.21	0.17	0.25
IV	0.43	0.09	0.75
BDS	0.24	0.21	0.28
additive	0.40	0.05	0.74

Table 3: Comparison of the estimated average treatment effect (ATE) for different subpopulations to OLS and IV estimates as well as the BDS matching estimates, and the additive model of Carneiro and Lee (2005).

estimate of BDS since matching is built on the assumption that conditional on observables, D is independent of the error term in the outcome equation. As covariates in the wage equation we used the variables from the final specification in Table 2, except for the father's interest in the education of the child.

Commonly, the linear IV estimate is interpreted as estimating the average treatment effect of those who are induced to attend college by the variables that are excluded from the outcome equation, see, e.g. the discussion in BDS and Imbens and Angrist (1994) as well as Card (2001). In our specification, following BDS, we have excluded the father's interest in the education of the child from the outcome equation. The estimate obtained from the additive model is close to our estimate for the population.

Notably, we estimate the average treatment effect to be lower for those who actually attend college. This difference can partly be explained by differences in observables since the average treatment effect depends on those observables in our model.

In general, all estimates which are obtained from a two stage (ours, OLS, IV, additive) procedure are relatively imprecise. We suppose that this is due to the first stage estimation error which is carried over into the second stage.

4.5. Average *Ceteris Paribus* Effects

Panel (1) in Table 4 contains estimates of average *ceteris paribus* effects and respective 95% confidence intervals. The set of covariates we included into the second step is the same as the one in the final specification for the first step, except that we leave out the father's interest in the education of the child. We also calculated estimates for alternative specifications, but the results did not change qualitatively.

The top rows contain estimates for $D = 0$ and the bottom rows for $D = 1$. Statistically significant determinants of wages are whether the father was professional, which resulted in a large wage increase both for $D = 0$ and even more so for $D = 1$ and the father's years of education, but only for $D = 0$. In general, our estimates were quite imprecise. Yet, as we have already seen above, this is also the case for the standard linear IV estimates. Therefore, we feel that this lack of precision is not a property of our estimation procedure, but a feature of our data.

	(1) average *ceteris paribus* effect			(2) test for unobserved heterogeneity			(3) bias in estimates when additivity is imposed		
	estimate	95% conf. int.		est. slope	95% conf. int.		bias	95% conf. int.	
				NO COLLEGE DEGREE					
ABILITY MEASURES									
math ability at 7	0.001	-0.001	0.003	0.000	-0.008	0.009	0.000	-0.002	0.002
math ability at 11	0.002	-0.002	0.006	0.000	-0.014	0.016	0.000	-0.006	0.003
verbal ability at 7	0.003	-0.002	0.007	0.007	-0.007	0.018	-0.002	-0.006	0.002
verbal ability at 11	0.000	-0.004	0.003	-0.001	-0.014	0.012	0.000	-0.005	0.002
SECONDARY SCHOOL TYPE RELATIVE TO COMPREHENSIVE SCHOOL									
Grammar	0.021	-0.230	0.275	0.099	-1.063	1.267	-0.003	-0.293	0.210
Private	0.159	-0.099	0.418	0.040	-1.260	1.295	-0.026	-0.362	0.158
SOCIAL CLASS OF THE FATHER RELATIVE TO UNSKILLED									
professional	0.293	0.007	0.757	-1.662	-3.386	-0.245	-0.316	-0.663	0.095
intermediate	-0.022	-0.116	0.087	-0.224	-0.680	0.257	0.017	-0.130	0.074
FAMILY BACKGROUND VARIABLES WHEN THE CHILD WAS 16									
father's years of education	0.034	0.007	0.064	-0.008	-0.145	0.114	-0.010	-0.041	0.016
missing	0.331	0.053	0.660	0.062	-1.305	1.358	-0.109	-0.441	0.165
mother was employed	0.023	-0.053	0.119	0.007	-0.328	0.426	-0.003	-0.035	0.137
number of siblings	-0.019	-0.053	0.011	-0.022	-0.126	0.084	0.006	-0.016	0.048
CONSTANT	1.269	0.566	2.014	-0.318	-1.970	1.368	0.253	-0.542	0.901
				COLLEGE DEGREE					
ABILITY MEASURES									
math ability at 7	-0.002	-0.010	0.005	-0.003	-0.011	0.006	0.003	-0.004	0.012
math ability at 11	0.004	-0.010	0.015	0.002	-0.012	0.015	-0.003	-0.014	0.012
verbal ability at 7	-0.011	-0.024	0.004	-0.010	-0.022	0.003	0.010	-0.004	0.024
verbal ability at 11	-0.013	-0.024	0.000	-0.014	-0.025	0.000	0.013	-0.003	0.021
SECONDARY SCHOOL TYPE RELATIVE TO COMPREHENSIVE SCHOOL									
Grammar	0.238	-0.396	0.758	0.303	-0.464	0.932	-0.259	-0.752	0.399
Private	0.528	-0.079	1.139	0.569	-0.197	1.331	-0.435	-1.048	0.170
SOCIAL CLASS OF THE FATHER RELATIVE TO UNSKILLED									
professional	0.794	0.054	1.540	1.035	0.080	2.090	-0.823	-1.611	-0.130
intermediate	0.174	-0.286	0.633	0.152	-0.364	0.686	-0.188	-0.721	0.200
FAMILY BACKGROUND VARIABLES WHEN THE CHILD WAS 16									
father's years of education	-0.006	-0.093	0.063	-0.014	-0.124	0.071	0.015	-0.046	0.110
missing	0.108	-0.873	0.911	0.006	-1.161	0.943	-0.012	-0.692	1.077
mother was employed	-0.001	-0.277	0.298	-0.075	-0.374	0.269	0.033	-0.253	0.323
number of siblings	-0.017	-0.123	0.074	0.003	-0.108	0.102	-0.004	-0.071	0.126
CONSTANT	4.126	2.578	5.583	1.998	0.474	3.147	-2.078	-3.821	-0.773

Table 4: Average *ceteris paribus* effects, test for unobserved heterogeneity, and estimates for the bias from imposing additivity.

Panel (2) contains the result of a test for unobserved heterogeneity. We say that unobserved heterogeneity is present whenever the impact of a component of X, including the constant, depends on V. Therefore, the null hypothesis is that the derivative of the conditional average *ceteris paribus* effect with respect to V is zero at all $V = v$. This implies that the linear approximation to the slope is zero. (2) contains estimated linear approximations to the slope of $\widehat{\mathbb{E}}[\varphi(d, U, v)]$, as well as bootstrapped confidence intervals. Notice that here, we face two sources of estimation error. First, the error that stems from estimating the conditional average *ceteris paribus* effect itself and second, the error from estimating the linear approximation to its slope. The presence of essential heterogeneity is significant at the 5% level if 0 lies outside the confidence interval. Using this test, we find evidence for essential heterogeneity in the impact of the father being professional for both $D = 0$ and $D = 1$ and overall for $D = 1$, *via* the constant term.[11]

This essential heterogeneity has the interpretation of a nonseparability between the effect of X and V on Y. It is a key advantage of the techniques developed in this paper to allow us to control for this nonseparability. In panel (3), we raise the question whether imposing the absence of this nonseparability, i.e. imposing the additive model of Carneiro, Heckman, and Vytlacil (2005) and Carneiro and Lee (2005), results in biases of average *ceteris paribus* effects. We report estimates of the bias that results from imposing separability. The estimates were obtained by comparing our estimator to a simple series estimator of the additive model $Y = \mu(D, U, V) + X'_{-1}\beta(D, U)$ in which the effect of X on Y is not allowed to depend on V. A cross validation yields that only a linear term in P should be included into the regression of Y on X conditional on D in order to calculate estimates of average *ceteris paribus* effects. Clearly, this proceeding is far less elaborate than the double-residual regression procedure that is carried out in, e.g., Carneiro, Heckman, and Vytlacil (2005) and Carneiro and Lee (2005). Therefore, we prefer to interpret our estimates of the biases only as rough estimates or first approximations. However, the results in panel (2) indicate already that the additive model is misspecified for our data so that it is not surprising that we estimate the bias to be significant for the impact of the father being professional and the constant term for $D = 1$.

4.6. Conditional Average *Ceteris Paribus* Effects and Sorting

Figure 3 and 4 in the Appendix contain estimates of conditional average *ceteris paribus* effects. They show the respective dependence of the impact of covariates on wages as a function of D and V. Notice that according to the selection model low values of V induce individuals to attend college so that we should think of *low* values of V as representing *high* unobservable ability. For example, whereas the impact of the father being professional on wages is increasing in unobservable ability for $D = 0$, it is decreasing for $D = 1$.

Since X varies across individuals, it is helpful to take a closer look at the dependence of the marginal treatment effect on V when X varies across individuals. Carneiro and Lee (2004, footnote 3) point out that individuals base their selection into educational on their *comparative advantage* with respect to monetary benefits if the marginal treatment effect is higher for those

[11]Carneiro, Heckman, and Vytlacil (2005) and Heckman, Urzua, and Vytlacil (2004) suggest to test for essential heterogeneity by checking whether the expected value of wages given P and X is linear in P by fitting polynomials to the data. Using this test, we were not able to reject the null of no essential heterogeneity for our data.

	fraction	95% conf. int.	
level, $D = 0$	0.62	0.04	0.93
level, $D = 1$	0.65	0.32	0.96
marginal treatment effect	0.54	0.47	0.60

Table 5: Fractions of observations for which the CASF (level) and the marginal treatment effect is increasing in V. Linear approximations to the slope were calculated.

individuals who go to college, i.e. if the marginal treatment effect is falling in V conditional on observables X.[12] Variation in covariates induces variation in the slope of the marginal treatment effect. Therefore, we estimated a linear approximation to the slope of the marginal treatment effect for every individual.

Table 5 contains the fractions of the population for which, respectively, the slope of the CASF and the marginal treatment effect are positive. In order to obtain those numbers, linear approximations to the slope were estimated. The numbers indicate that the way wages depend on what we labelled unobserved ability, V, is nontrivial.

As for the slope of the levels, the slope is positive in about 60 per cent of the cases. A positive slope implies that

$$x'\mathbb{E}[\varphi(0, U, V)|D = 1] < x'\mathbb{E}[\varphi(0, U, V)|D = 0]$$
$$x'\mathbb{E}[\varphi(1, U, V)|D = 1] < x'\mathbb{E}[\varphi(1, U, V)|D = 0].$$

Hence, the numbers indicate that in about 60 per cent of the cases those who actually graduated from college earn less compared to what those, who did not graduate from college, would earn, had they been forced to do so. Conversely, those who did not go to college earn more than those who did go to college would have earned, had they been prevented from doing so. This is in line with our earlier finding that treatment effects are higher for college non-graduates compared to college graduates. However, notice that this is only an analysis of monetary benefits, neglecting the costs of attending college which could have been prohibitively high for those who did not in fact attend college.

Surprisingly, only for about 46 per cent of the individuals the slope of the marginal treatment effect is negative. Hence, the comparative advantage hypothesis is only supported for these 46 per cent of the individuals. For about 54 per cent of the individuals, the slope is positive. This is in contrast to the findings in previous studies including Willis and Rosen (1979) and Carneiro and Lee (2004). One explanation could be that both of these studies do not allow the effect of V on wages to depend on X. In fact, as we have seen in Table 3, such estimates would be biased for our data.

We shall end with the conjecture that the comparative advantage hypothesis, which is a

[12]See, e.g., Roy (1951) for the impact of selection of individuals based on their comparative advantage on the income distribution, Sattinger (1978) for an empirical study of respective comparative advantages of individuals in the performance of tasks, Willis and Rosen (1979) for a parametric study of the returns to college education in the presence of such selection, as well as Carneiro and Lee (2004) for a semiparametric analysis. Heckman and Sedlacek (1985, 1990) develop models of the sectoral allocation of workers based on comparative advantage.

central concept in Economics, could well be reconciled with these findings once nonmonetary costs and benefits are included in the analysis. Just to give an example, it could well be that a college degree is associated with nonpecuniary benefits such as the pleasure of being educated which represent an *additional* return that has not been focussed on in this study. Clearly, such nonmonetary costs and benefits might again well be correlated with unobserved ability, family background, and social class. After all, we understand our results as evidence for nontrivial sorting patterns that are not solely based on monetary considerations. Therefore, we strongly believe that more research, and other data, are of need in order to better understand the sorting patterns into educational levels.

5. Concluding Remarks

In this paper, we have proposed and implemented a semiparametric estimator for expected wage levels and their dependence on the endogenous schooling choice.

The virtue of our approach to the problem lies in dimensionality reduction along the dimension of the usually higher dimensional vector of exogenous covariates. Moreover, we are able to circumvent the problem of limited support of the propensity score given the vector of covariates since we require only conditions on the unconditional support of P to hold. At the same time, we do not impose any limiting restrictions on the joint distribution of unobservables.

The estimator we propose is a two step version of a local linear regression estimator. The usefulness of our approach was shown in turn of the empirical analysis. In particular, our results suggest that differences in wages can be attributed to differences in observables in interaction with unobserved ability. In previous studies, e.g. by Carneiro, Heckman, and Vytlacil (2005) and Carneiro and Lee (2005), this complementarity between observables and unobservables was largely neglected for reasons of tractability. In this paper, we have suggested an estimation procedure which does allow for such effects on the one hand and which is easily implementable on the other.

The results of the empirical analysis are manyfold. First, we find that measured ability, social class, secondary school type, and family background have explanatory power for the decision to attend college. Second, with an estimate of 0.46 for the population, we find the monetary return to college education to be sizable with returns for college graduates being lower than for college non-graduates. Third, our estimates do not support the hypothesis of sorting into schooling based on comparative advantage with respect to the monetary returns. Forth and last, we find nonseparabilities between the impact of observables, e.g. whether the father is professional, and unobserved ability on wages and show that biases arise once an additive structure is imposed. We feel that this shows the usefulness of our approach.

Appendix

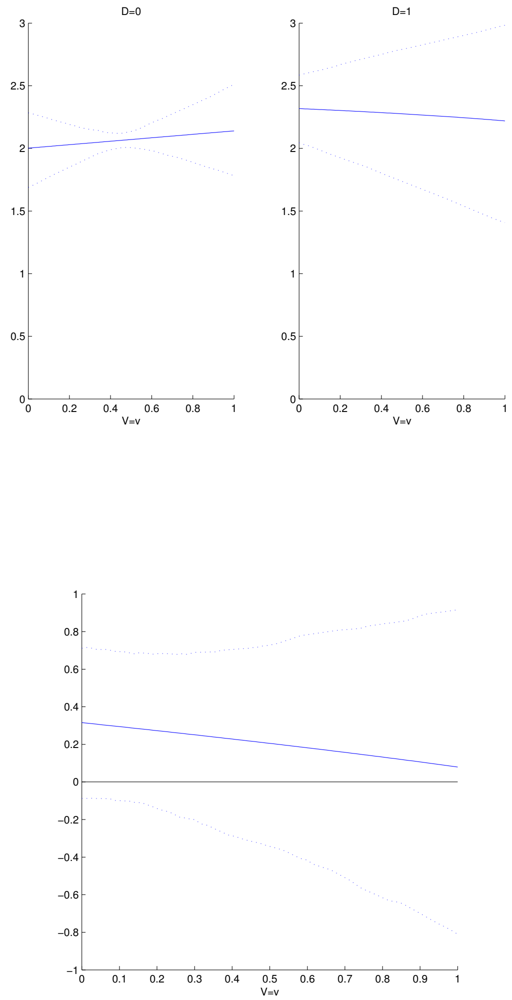

Figure 2: Point estimates and bootstrapped 95% confidence intervals of the conditional average structural function (top) and the marginal treatment effect (bottom). Reported for a representative individual with median characteristics.

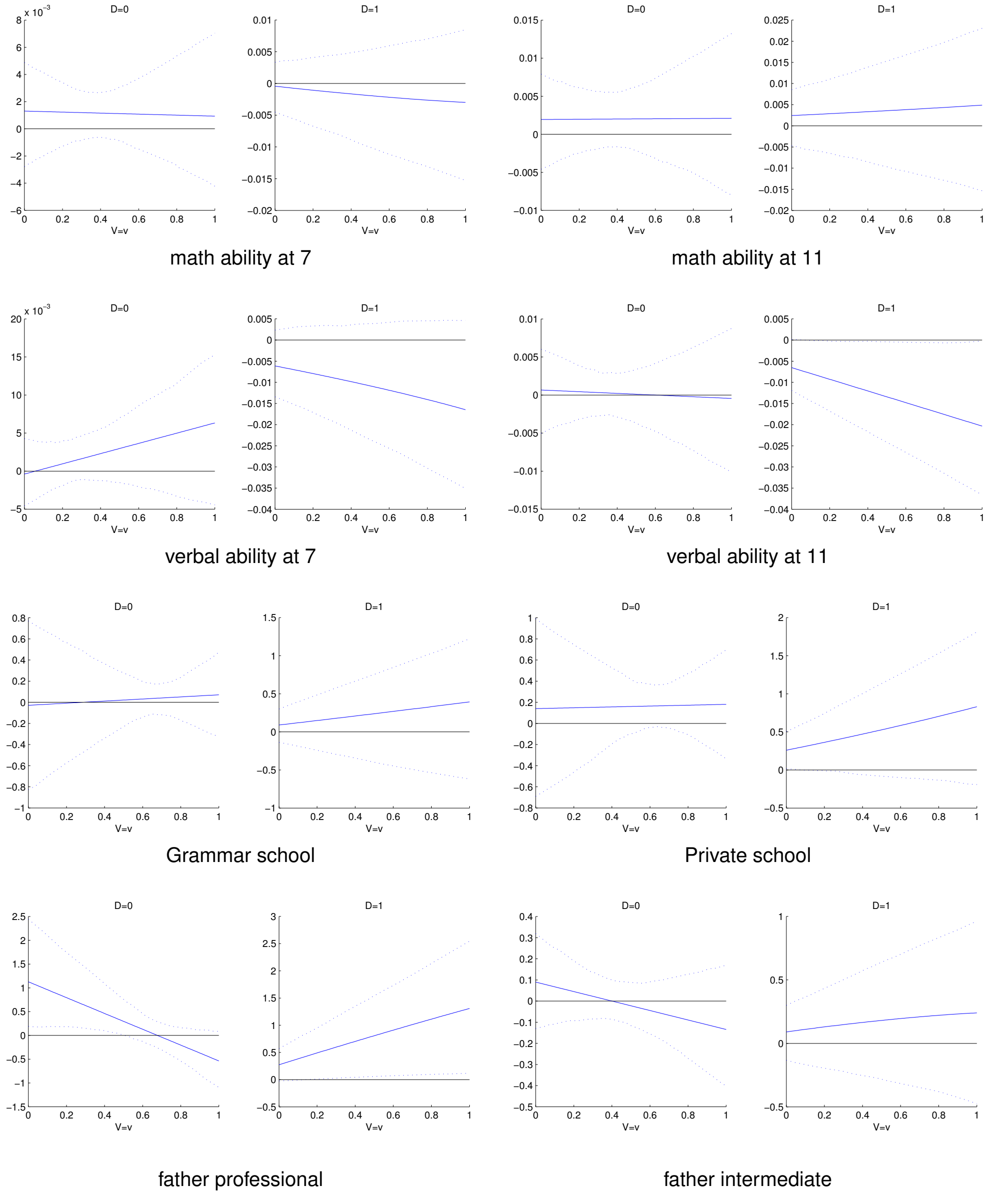

Figure 3: Conditional average *ceteris paribus* effects 1/2. Point estimates and bootstrapped 95% confidence intervals.

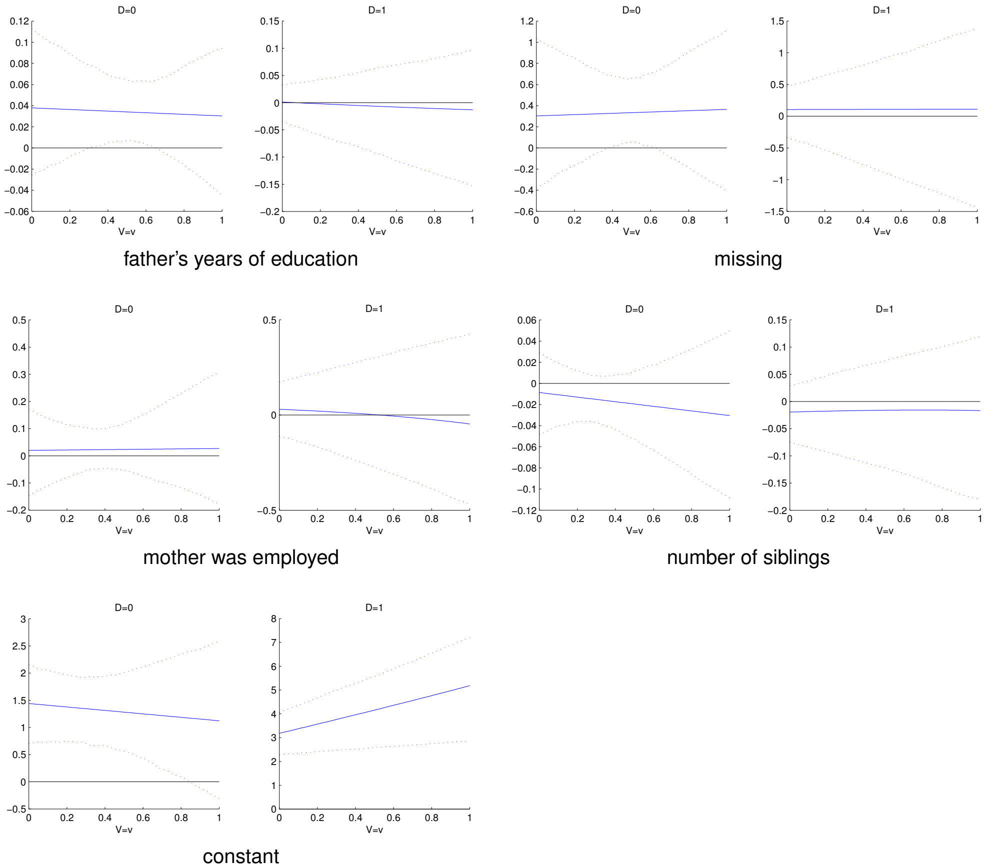

Figure 4: Conditional average *ceteris paribus* effects 2/2. Point estimates and bootstrapped 95% confidence intervals.

References

ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70(1), 91–117.

ANGRIST, J. D., K. GRADDY, AND G. W. IMBENS (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies*, 67(3), 499–527.

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

——— (2001): "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 15(4), 69–85.

ASHENFELTER, O., AND J. D. MOONEY (1968): "Graduate Education, Ability, and Earnings," *Review of Economics and Statistics*, 50(1), 78–86.

BECKER, G. S. (1967): "Human Capital and the Personal Distribution of Earnings," Woytinski Lecture No. 1, Ann Arbor: University of Michigan Press.

BJÖRKLUND, A., AND R. MOFFITT (1987): "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics*, 69(1), 42–49.

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(3), 473–512.

BLUNDELL, R., AND J. L. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics*, ed. by L. Hansen, Amsterdam. North Holland.

CARD, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127–1160.

CARNEIRO, P., J. J. HECKMAN, AND E. VYTLACIL (2005): "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education," Mimeograph.

CARNEIRO, P., AND S. LEE (2004): "Comparative Advantage and Schooling," Mimeograph.

——— (2005): "Ability, Sorting and Wage Inequality," c*e*mmap Working Paper CWP 16/05, University College London.

CHAMBERLAIN, G. (1977): "Education, Income, and Ability Revisited," *Journal of Econometrics*, 5, 241–259.

CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–261.

CHERNOZHUKOV, V., G. W. IMBENS, AND W. K. NEWEY (2004): "Instrumental Variable Identification and Estimation of Nonseparable Models via Quantile Conditions," MIT Working Paper.

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71(5), 1405–1441.

——— (2005): "Nonparametric Identification Under Discrete Variation," *Econometrica*, 73(5), 1525–1550.

CLEVELAND, W. S., E. GROSSE, AND W. M. SHYU (1991): "Local Regression Models," in *Statistical Models in S*, ed. by J. M. Chambers, and T. J. Hastie, pp. 309–376. Wadsworth/Brooks-Cole, Pacific Grove, CA.

DAROLLES, S., J.-P. FLORENS, AND E. RENAULT (2003): "Nonparametric Instrumental Regression," IDEI Working Paper 228.

DAS, M. (2005): "Instrumental Variables Estimation of Nonparametric Models with Discrete Endogenous Regressors," *Journal of Econometrics*, 124(2), 335–361.

FAN, J., AND W. ZHANG (1999): "Statistical Estimation in Varying Coefficient Models," *Annals of Statistics*, 27(5), 1491–1518.

FISHER, R. A. (1935): *The Design of Experiments*. Oliver & Boyd, Edinburgh.

GAREN, J. (1984): "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable," *Econometrica*, 52(5), 1199–1218.

GOLDBERGER, A. S. (1972): "Structural Equation Methods in the Social Sciences," *Econometrica*, 40(6), 979–1001.

——— (1989): "Economic and Mechanical Models of Intergenerational Transmission," *American Economic Review*, 79(3), 504–513.

GRILICHES, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45(1), 1–22.

GRILICHES, Z., AND W. M. MASON (1972): "Education, Income, and Ability," *Journal of Political Economy*, 80(3, Part 2: Investment in Education: The Equity-Efficiency Quandary), S74–S103.

HANSEN, W. L., B. A. WEISBROD, AND W. J. SCANLON (1970): "Schooling and Earnings of Low Achievers," *American Economic Review*, 60(3), 409–418.

HASTIE, T., AND R. TIBSHIRANI (1993): "Varying-Coefficient Models," *Journal of the Royal Statistical Society Series B (Methodological)*, 55(4), 757–796.

HAVEMAN, R., AND B. WOLFE (1995): "The Determinants of Children's Attainments: A Review of Methods and Findings," *Journal of Economic Literature*, 33(4), 1829–1878.

HECKMAN, J. J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 49, 931–959.

HECKMAN, J. J., AND R. ROBB (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman, and B. Singer, chap. 4, pp. 159–244. Cambridge Univeristy Press, Cambridge, MA.

——— (1986): "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in *Drawing Inference from Self-Selected Samples*, ed. by H. Wainer, pp. 61–113. Springer, Heidelberg, New York, with discussion.

HECKMAN, J. J., AND G. SEDLACEK (1985): "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy*, 93(6), 1077–1125.

——— (1990): "Self-Selection and the Distribution of Hourly Wages," *Journal of Labor Economics*, 8(1, Part 2: Essays in Honor of Albert Rees), S329–S363.

HECKMAN, J. J., S. URZUA, AND E. VYTLACIL (2004): "Understanding Instrumental Variables in Models with Essential Heterogeneity," Memeograph.

HECKMAN, J. J., AND E. J. VYTLACIL (1998): "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling," *Journal of Human Resources*, 33(4), 974–987.

——— (1999): "Local Instrumental Variables and Latent Variables Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96, 4730–4734.

——— (2000a): "Local Instrumental Variables," in *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, ed. by C. Hsiao, K. Morimune, and J. Powell, pp. 1–46. Cambridge University Press, Cambridge.

——— (2000b): "Local Instrumental Variables," NBER Working Paper No. T0252.

——— (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

IMBENS, G. W., AND W. K. NEWEY (2003): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," MIT Working Paper, Presented at the 2003 EC2 conference held in London.

KELEJIAN, H. H. (1971): "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables," *Journal of the American Statistical Association*, 66(334), 373–374.

KLEIN, T. J. (2006): "Heterogeneous Treatment Effects: Instrumental Variables without Monotonicity?," Mimeograph.

LEIBOWITZ, A. (1974): "Home Investments in Children," *Journal of Political Economy*, 82(2, Part 2: Marriage, Family Human Capital, and Fertility), S111–S131.

NEWEY, W., AND J. L. POWELL (2003): "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*, 71(5), 1565–1578.

NEWEY, W., J. L. POWELL, AND F. VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67(3), 565–603.

NEWEY, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimates," *Journal of Econometrics*, 79, 147–168.

PINSKE, J. (2000): "Nonparametric Regression Estimation Using Weak Separability," University of British Columbia.

ROBINSON, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56(4), 931–954.

ROSENBAUM, P. R., AND D. B. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70(1), 41–55.

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

SATTINGER, M. (1978): "Comparative Advantage in Individuals," *Review of Economics and Statistics*, 60(2), 259–267.

SOLMON, L. C., AND P. J. TAUBMAN (1973): *Does College Matter? Some Evidence of the Impacts of Higher Education*. Academic Press, New York.

TAUBMAN, P. J. (1973): "Schooling, Ability, Non Pecuniary Rewards, Socioeconomic Background and the Lifetime Distribution of Earnings," in *The Distribution of Economic Well-Being*, ed. by T. F. Juster, pp. 419–510. Ballinger Publishing Company, Cambridge.

TAUBMAN, P. J., AND T. J. WALES (1972): "Mental Ability and Higher Educational Attainment in the Twentieth Century," NBER and Carnegie Commission on Higher Education.

VYTLACIL, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70(1), 331–341.

WEISBROD, B. A., AND P. KARPOFF (1968): "Monetary Returns to College Education, Student Ability, and College Quality," *Review of Economics and Statistics*, 50(4), 491–497.

WILLIS, R. J., AND S. ROSEN (1979): "Education and Self-Selection," *Journal of Political Economy*, 87(5, Part 2: Education and Income Distribution), S7–S36.

XIA, Y., AND W. K. LI (1999): "On the Estimation and Testing of Functional-Coefficient Linear Models," *Statistica Sinica*, 9, 735–757.

YATES, F. (1937): "The Design and Analysis of Factorial Experiments," Discussion paper, Technical Communication no. 35 of the Commonwealth Bureau of Soils.