Estimating Conditional Average Structural Functions in Nonadditive Models with Binary Endogenous Variables*

Tobias J. Klein†

preliminary version: May 31, 2005

Abstract

We propose an estimator for identifiable features of correlated random coefficient models with binary endogenous variables and nonadditive errors in the outcome equation. These 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 instrumental variables. In an application, we estimate expected levels of wages as well as a variety of average *ceteris paribus* effects of changes in covariates and schooling. Moreover, we uncover the dependencies between wages and unobserved ability, measured ability, social background, and college education.

*I am especially grateful to Richard Blundell, Pedro Carneiro, Andrew Chesher, Stefan Hoderlein, Enno Mammen, and Melanie Schienle 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.

†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

Consider a correlated random coefficient model of the form

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

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

with a binary endogenous regressor D such as an indicator for college attendance in the returns to schooling case or an indicator for program participation in policy evaluation. Y is an outcome such as wages, X is a K-vector of covariates in the outcome equation (1), and Z is a vector of covariates in the selection equation (2), which we shall refer to as a vector of instrumental variables. X and Z include a constant as their respective first elements. U is a vector and V is a scalar. We observe Y, D, X, and Z and assume that (U, V) are jointly independent of (X, Z) and that Z is independent of V.

This correlated random coefficients model is similar to the one discussed by Heckman and Vytlacil (1998). However, we allow both the random coefficient and D to depend on V. Hence, we do not exclude selection on unobservables of the type discussed by Heckman and Robb (1985, 1986).[1]

The model 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 (forthcoming) shows that in this case set identification may still be feasible.

Here, we are not interested in identification of the structure itself, i.e. the coefficient function $\varphi(D, U, V)$, but in identifiable features such as the expected level of Y for a given D, X, and V,[2]

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

We will refer to this as the *conditional average structural function*. This terminology was introduced by Blundell and Powell (2001) who suggest to focus on the average structural function, $\mathbb{E}[Y|D = d, X = x]$, and argue that it is a "parameter of central interest in the analysis of semiparametric and nonparametric models with endogenous regressors".[3] We believe that the dependence of the average structural function on V is of central economic interest and can be given a structural interpretation. For example, if Y is earnings and D is an indicator for college, V has the interpretation of unobserved ability. Then, high values of V are associated with low unobserved ability.

[1] This selection, together with selection on observables Z, gives rise to the endogeneity of D. In the statistics literature, V is sometimes referred to as a confounding variable, see e.g. Fisher (1935, Ch. 7) and Yates (1937).

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

[3] Imbens and Newey (2003) call it the average conditional response.

Likewise, a second object of interest is the expected *ceteris paribus* effect of changes in X_k for a given D, X_{-k}, and V,[4]

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

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

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

This is Björklund and Moffit's (1987) marginal treatment effect. In the returns to schooling example, since high values of V have the interpretation of low unobservable ability, we would expect the marginal treatment effect to be nonincreasing in V.[5]

Our approach combines the identification strategy suggested by Heckman and Vytlacil (1999, 2000a, 2000b) and desirable features of instrumental variables estimators in a natural way. In particular, we suggest to estimate identifiable features of the outcome equation by a local linear smoothing estimator. It is built on the conventional two stage least squares IV estimator, except that we let the coefficients depend on the value of $P(Z)$. It is easily implementable and does not rely on strong support conditions.[6]

First applications building on the identification result of Heckman and Vytlacil are Carneiro, Heckman, and Vytlacil (2003) and Carneiro and Lee (2004). However, the model they estimate is restrictive. 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 estimate an outcome equation of the form

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

That is, they do not allow for the effect of X on Y to depend on V. This dependence is an important aspect of unobserved heterogeneity and—as we have argued before—is of economic interest in many applications with binary endogenous variables.

In our application, we implement the proposed estimator using U.K. data. The virtue of the data set that was used, the National Child Development Survey (NCDS), is that

[4]The kth element of a vector x is denoted by x_k. The remaining elements are denoted by x_{-k}.

[5]The well known average treatment effect for a given $X = x$ is given by

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

where $f_V(\cdot)$ is the density of V.

[6]Heckman and Vytlacil (2005) discuss the relationship between the conventional IV estimator and the "local" IV estimator of Heckman and Vytlacil (1999, 2000a, 2000b). Nonparametric identification requires conditions on the support of $P(Z)$ conditional on X to hold. In our approach, these conditions have only to hold for the unconditional support of $P(Z)$. This will be discussed in detail in Section 2.

detailed ability measures and family background variables are available. In particular, we seek to identify the determinants of earnings. Hereby, we focus on the differences in earnings that are due to a higher education degree, social background as well as the unobserved ability in V.

The paper is organized as follows. In Section 2, we present the econometric model, the identification result, and the proposed estimator. Section 3 contains the empirical application and Section 4 concludes. A description of the data including summary statistics as well as detailed results of the econometric analysis are contained in an appendix.

2. ECONOMETRIC MODEL

Our point of departure is the correlated random coefficients model that was given in (1) and (2). We restate it for convenience

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

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

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 stronger than the usual instrumental variables condition that $(U, V) \perp\!\!\!\perp Z | X$ needed for nonparametric identification. However, it is weaker than the exclusion restrictions usually invoked in correlated random coefficient models.[7] 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.[8]

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.*

[7] A version of the assumptions made in Heckman and Vytlacil (1998) is that V in (1) can be replaced by the elements in Z which are not in X.

[8] 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 only be identified up to normalizations on the joint distribution of unobservables. 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.

Assumption 2(i) ensures that all parameters of interest are well defined. Assumption 2(ii) implies that V is a continuous random variable. Then, w.l.o.g., we can normalize V to be uniformly distributed, see, e.g., Vytlacil (2002) for details. Then, 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.

In the next subsection, we show identification under Assumption 1 and 2. Estimation built on local linear smoothing is dealt with thereafter.

2.1. Identification

In this subsection, we show that, under appropriate support and differentiability conditions, the parameters of interest can be point identified from observations.

From the model in (1), it follows that

$$\mathbb{E}[Y|X = x, P = p, D = 1] = \mathbb{E}[X'\varphi(1, U, V)|X = x, P = p, D = 1] \tag{3}$$

which is equal to

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

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

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

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 (3) is identified from observations at points $X = x$ and $P = p$, $\beta(1, p)$ is identified if we observe at least K linearly independent values of X for every D (rank condition).

Starting from this, we can show that the parameters of interest are identified. We state the result in a theorem which resembles Lemma 1 from Carneiro and Lee (2004).

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 functions of P are identified from observations.

THEOREM 1 (Carneiro and Lee): *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 conditional average structural function is identified at $V = p$, where p is a limit point of the support of P, and given by*

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

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

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

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. □

From Theorem 1 it follows immediately that average *ceteris paribus* effects are identified. The marginal treatment effect is identified if p is a limit point of the support of P for both $D=0$ and $D=1$.

2.2. 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 the propensity score $P(Z)$. 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 on 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 $\hat{p}$ about $p=\hat{p}$ which yields

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

where $\tilde{p}$ is a point between p and $\hat{p}$. 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 sense. However, confidence intervals, accounting for the first step estimation error, were obtained using a bootstrap procedure.[9]

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. The next section contains the empirical application.

3. Returns to College Education in the U.K.

The question of how to estimate the returns to schooling and college education is one of the classical questions in econometrics for several reasons.[10]

First, it is of high practical relevance. Earnings, interpreted as a function of ability and education, are market prices which, in turn, are indicators of scarcity of resources in an economy. Therefore, the returns to college education are an indicator for the scarcity of college graduates relative to high school graduates and the characterization of returns to college education are a valuable piece of information for politicians designing educational policy.[11]

Second, it is still an open question whether we should think of college education and schooling as an investment in human capital or as a signalling device.[12] Human capital theory assumes that a person's ability and time spent in educational institutions enter as production factors into the production of human capital.[13] Then, it can be argued that

[9]Fan and Zhang (1999) show that if the degree of smoothness is different across coefficient functions the rate of convergence of the estimator will in general not be optimal. They suggest a two step procedure to overcome this problem.

[10]For two excellent surveys of the literature on the returns to schooling see Griliches (1977) and Card (2001). For early surveys on the returns to college education see Solmon and Taubman (1973) and Taubman and Wales (1974).

[11]If college education is complementary to a student's unobserved and observed ability, these returns are likely to be heterogeneous across students just because of differences in ability.

[12]See Becker (1964, 1993) on human capital theory and, e.g., Taubman and Wales (1973), Spence (1972), Stiglitz (1973) and Arrow (1974) on signalling.

[13]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,

the higher a person's human capital the more she will earn since her productivity will increase in the amount of human capital he has acquired, and this will be rewarded by the labor market. On the other hand, once we interpret educational institutions as signalling devices, education by itself does not increase a person's productivity but reveals it. For these reasons, it is consistent with both signalling and human capital theory that ability and productivity are interrelated. Furthermore, since ability is hardly measurable, or—in econometric terms—is likely to be partly unobservable or measured with error[14], it is in general not possible to reject one of the two theories in favor of the other.

Third, 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.[15] A variety of approaches to this challenging problem has been taken over the last four decades. Identifying assumptions include parametric assumptions, conditional (mean) independence and monotonicity in order to identify mean returns. Also, quantile invariance has proved to be a powerful identifying assumption.[16] However, most of these approaches rely on the presence of instrumental variables that can be excluded from the earnings equation. Instrumental variables that have been used are quarter of birth (Angrist and Krueger 1991) and parental interst in education (Blundell, Dearden, and Sianesi forthcoming) 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 suppose the selection model in (2),

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

was appropriate. Then, for a given vector of observables Z, and hence for a given $P = P(Z)$, only those with values of V less than or equal to P would decide to attend college.

p. 8) suggests that one could interpret ability also as "initial human capital". 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. For the link between ability and earnings see Ashenfelter and Mooney (1968), Griliches and Mason (1972), Hansen, Weisbrod, and Scanlon (1970), Weisbrod and Karpoff (1968) and Leibowitz (1974).

[14]Griliches (1977) discusses econometric consequences when ability is measured with error. Alternatively, we could think of unobserved ability as being a left-out variable (Chamberlain 1977, e.g.).

[15]See, e.g., Heckman (1978), Willis and Rosen (1979), and Garen (1984).

[16]For distributional assumptions see, e.g., Heckman (1978), and Aakvik, Heckman, and Vytlacil (forthcoming). Conditional independence is assumed in Rosenbaum and Rubin (1983). Heckman and Vytlacil (1998) exploit additivity of the error term in a random coefficient framework. Garen (1984), Heckman (1978), Newey, Powell, and Vella (1999) as well as Pinske (2000) and Blundell and Powell (2003) persue a control function approach based on a similar structure. 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. Imbens and Angrist (1994), Angrist, Graddy, and Imbens (2000), Carneiro, Heckman, and Vytlacil (2003) and Heckman and Vytlacil (2005) as well as Abadie, Angrist, and Imbens (2002) exploit monotonicity. Quantile invariance is relied on in Chernozhukov, Imbens, and Newey (2004) and Chernozhukov and Hansen (2005).

Therefore, *low* values of V have the interpretation of *high* unobserved ability.[17] Hence, *a priori*, we would expect the conditional average structural function to be nonincreasing in V and the average *ceteris paribus* effects to depend on V.

We estimate the expected levels of earnings for college graduates and nongraduates conditional on a vector of covariates containing information on social background as well as other characteristics.[18]

For a detailed data description the reader is referred to the appendix and Blundell, Dearden, and Sianesi (forthcoming). Tables and figures referred to here can also be found in the appendix. Table 1 contains a list of variables and summary statistics for our data. Table 2 contains OLS coefficients from a regression of log wages on a dummy for college and the covariates in our data. This regression can be interpreted as a best linear predictor under a square loss, i.e. a descriptive statistic. The covariates include ability test scores, school type, family background, and region. The point estimates suggest that ability has a significant impact on earnings. Figure 1 plots the distribution of ability test scores at the age of 7 and 11, respectively.

We estimate a probit model in order to obtain fitted probability values of college attendance for every individual.[19] Following Blundell, Dearden, and Sianesi (forthcoming) we argue that parental interest in education can be excluded from the outcome equation. These instrumental variables generate some variation in the fitted values even conditional on X. Note, however, that variation is not even needed conditional on X for identification in our model—as was already pointed out above. Figure 2 shows the unconditional support of P for observations with $D = 0$ and $D = 1$, respectively. In Sections 1 and 2 the fitted values have been referred to as values of the propensity score. Note that whereas the interpretation of the estimated probit coefficients as causal *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.[20] The results from the reduced form probit model are contained in Table 3.[21] Figure 2 shows the support of the propensity score. For both $D = 0$ and $D = 1$ it is almost equal to the full unit interval. Note that the distributions differ between $D = 0$ and $D = 1$. This indicates that the variables included in Z have explanatory power. Figure 3 is a scatter plot of log hourly wages against values of the propensity score.

---

[17]We interpret unobservable ability as being the confounding factor, cf. footnote 1. In principle, V is the projection of all unobservable factors that give rise to the endogeneity of D into a scalar. Examples of these factors include unobservable wealth, unobservable status or social origin, and ideosyncratic preferences.

[18]To be precise, we distinguish between higher education graduates ($D = 1$) and those who have obtained just A-levels ($D = 0$).

[19]In principle, we could also have estimated these fitted values by ordinary least squares, see Kelejian (1971) and the discussion in Angrist and Krueger (2001).

[20]Willis and Rosen (1979) use longitudinal data which allows them to estimate both a reduced form and a structural probit. See also the discussion in Angrist and Krueger (2001) for an opposite point of view.

[21]For expositional purposes, we report them for the case in which we include dummies for quintiles of ability test scores as regressors. In the specification that was finally used for the first step estimates we used absolute values of these ability test scores.

The bandwidths for the second step estimates were chosen using a cross validation procedure. The criterion, the mean integrated square error, is shown in Figure 4. Our point estimates of the conditional average structural function as well as derivatives and differences thereof, including bootstrapped 95% pointwise confidence intervals, were obtained using the results from Section 2 (1,000 bootstrap repetitions). The estimates are shown in Figures 5 through 7. In our bootstrap procedure we acknowledge the fact that the propensity score is estimated in a first step by estimating it within every bootstrap repetition. Figure 8 contains the result of a simulation in which we generate the distribution of gross gains from college attendance for the whole population, the subset that attended college, and the subset that did not, respectively. That is, the distribution of marginal treatment effects of the treated, untreated, and both. Every cumulative distribution function reflects differences of individuals with respect to observables X, which we observe in our data, and unobservables V, which we randomly draw. To be specific, for individuals with $D = 0$, we draw values of V which are strictly greater than p from a uniform distribution. Similarly, for individuals with $D = 1$, we draw values of V which lie between 0 and p.

Figures 5 and 6 reveal that, for a representative individual with $X = x$ (see the Appendix for details), the gains from a college degree do not depend on unobserved ability. This is, we find that in contrast to the findings in Carneiro and Lee (2004) and Willis and Rosen (1979) our results do not support the comparative advantage hypothesis. However, Figure 7 shows that the effect of X on Y depends on the value of the unobserved ability V. In particular, Figure 7 suggests that high values of measured math ability at the age of 7 and 11 are associated with large premia for college graduates. That is, for lower values of observed math ability, expected earnings for our representative individual would have been decreasing in V. In general, the effects of observables on wages seem to depend on the confounding factor V. This shows that it was important to allow for dependence of the correlated random coefficient on V in our model. Otherwise, these effects would have been attributed either to the returns to college, which is likely if conventional instrumental variables methods are used (see Heckman and Vytlacil 2005), or to the effect of college education on earnings once only these effects are allowed to depend on V.

In the population, the observed difference in earnings can be traced back to a selection effect and a causal effect of a higher education degree. Figure 8 indicates that the benefits from a higher education degree are highest for those who did in fact not attend college. According to the model this can be traced back to differences in observables and unobservable ability. Still, this finding may be well in accordance with a desirable or even efficient allocation of college education once mostly high opportunity costs determine individuals' choice rather than financial limitations or social background.

4. Concluding Remarks

In this paper, we have proposed a set of assumptions which enabled us to derive a semiparametric estimator for the conditional average structural function which is of central interest to economists whenever we face the problem of potentially endogenous binary regressors. 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. On the other hand, 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 outlined in the empirical example, the returns to schooling in the United Kingdom. In particular, our results suggest that differences in wages can be attributed to differences in observables in interaction with unobserved ability. In previous studies, for example by Carneiro, Heckman, and Vytlacil (2003) and Carneiro and Lee (2004), 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.

Appendix

The National Child Development Survey (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 3rd and 9th March, 1958. Data was collected in 1965 (when members were aged 7 years), in 1969 (age 11), in 1974 (age 16), in 1981 (age 23), in 1985 (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.

The purpose of this paper is to identify and quantify the determinants of earnings in the United Kingdom. Following Blundell, Dearden, and Sianesi (forthcoming), our outcome of interest is log hourly wages in 1985, 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. Additionally, we say so for a variety of similar degrees, see Blundell, Dearden, and Sianesi (forthcoming) for details. We proceed similarly for a higher education degree. As they do, we focus on males. Detailed summary statistics are given in Table 1.

All results are reported for white males who attended Comprehensive school at the age of 16, whose father and mother had 9.4 and 9.6 years of education and are 44.5 and 42.4 years old, and whose father was an intermediate employee when the child was 16. At the age of 16, they lived in the London area, had 2 siblings and their mother was employed.

Tables and Figures

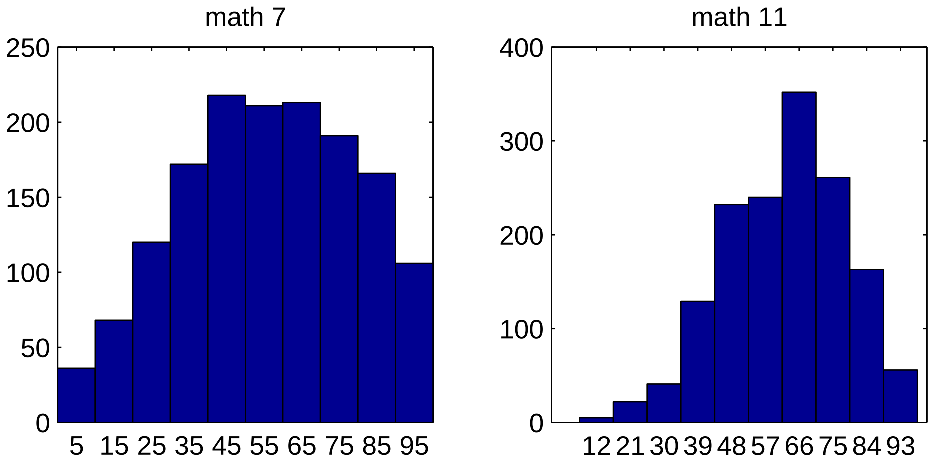

Figure 1: Distribution of math test scores.

	mean	*standard deviation*		*mean*	*standard deviation*
log hourly wage	2.186	0.414	mother's years of education	7.787	4.685
fraction of HE graduates	0.514	0.500	missing	0.243	0.429
white	0.978	0.146	father's years of education	7.836	5.021
math ability at 7			missing	0.255	0.436
lowest quintile	0.082	0.274	mother's age	41.612	10.753
second quintile	0.135	0.341	missing	0.048	0.214
third quintine	0.186	0.389	father's age	43.362	13.490
fourth quintile	0.204	0.403	missing	0.072	0.259
highest quintile	0.282	0.450	*father's social class when the child was 16*		
reading ability at 7			professional	0.070	0.256
lowest quintile	0.088	0.283	intermediate	0.185	0.389
second quintile	0.147	0.354	skilled non-manual	0.089	0.284
third quintile	0.198	0.399	skilled manual	0.273	0.445
fourth quintile	0.228	0.419	semi-skilled non-manual	0.006	0.074
highest quintile	0.228	0.419	semi-skilled manual	0.075	0.264
ability at 7 missing	0.112	0.315	unskilled	0.206	0.404
math ability at 11			missing	0.097	0.296
lowest quintile	0.048	0.213	*mother's interest in education*		
second quintile	0.115	0.319	expects too much	0.039	0.194
third quintile	0.152	0.359	is very interested	0.418	0.493
fourth quintile	0.223	0.416	has some interest	0.330	0.470
highest quintile	0.287	0.453	*father's interest in education*		
reading ability at 11			expects too much	0.018	0.132
lowest quintile	0.057	0.231	is very interested	0.323	0.468
second quintile	0.127	0.334	has some interest	0.216	0.412
third quintile	0.164	0.370	parental interest missing	0.080	0.271
fourth quintile	0.218	0.413	*region at the age of 16*		
highest quintile	0.259	0.438	North Western	0.106	0.308
ability at 11 missing	0.176	0.381	North	0.066	0.248
school type at the age of 16			East and West Riding	0.071	0.257
Comprehensive	0.445	0.497	North Midlands	0.073	0.260
Secondary	0.124	0.330	Eastern	0.070	0.256
Grammar	0.143	0.350	London and South East	0.148	0.355
Private	0.083	0.277	Southern	0.067	0.250
other	0.014	0.118	South Western	0.070	0.256
missing school type	0.191	0.393	Midlands	0.078	0.269
family background variables			Wales	0.052	0.221
mother was employed	0.533	0.499	other	0.103	0.304
number of siblings	1.517	1.503			

Summary statistics for all $N = 1977$ employed males who completed A-levels.

Table 1: Summary statistics.

covariate	*point estimate*	*standard error*	*t-statistic*
higher education	0.227	0.018	12.371
math ability at 7 relative to the lowest			
second quintile	-0.022	0.038	-0.590
third quintile	-0.019	0.038	-0.503
fourth quintile	0.007	0.037	0.189
highest quintile	0.058	0.038	1.543
reading ability at 7 relative to the lowest			
second quintile	0.093	0.038	2.485
third quintile	0.092	0.038	2.442
fourth quintile	0.095	0.039	2.444
highest quintile	0.077	0.040	1.927
ability at 7 missing	0.127	0.047	2.710
math ability at 11 relative to the lowest			
second quintile	0.005	0.049	0.100
third quintile	0.030	0.051	0.587
fourth quintile	0.044	0.052	0.850
highest quintile	0.059	0.054	1.083
reading ability at 11 relative to the lowest			
second quintile	0.026	0.046	0.572
third quintile	0.066	0.049	1.251
fourth quintile	0.102	0.052	1.969
highest quintile	0.092	0.055	1.677
ability at 11 missing	0.070	0.052	1.362
number of siblings	-0.018	0.007	-2.664

The dependent variable is log hourly wage at the age of 33. We also control for school type, family background, and region. $N = 1977$, $R^2 = 0.205$.

Table 2: OLS estimates.

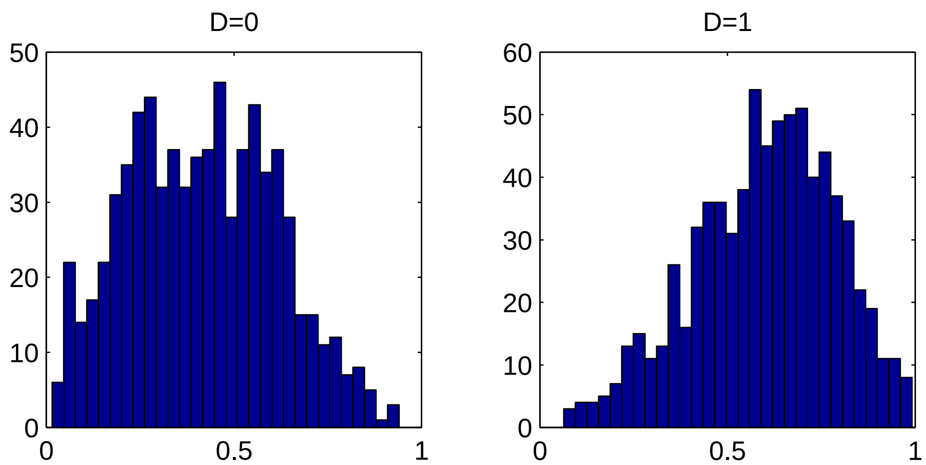

Figure 2: Distribution of values of the propensity score.

covariate	*dF/dx*	*standard error*	*t-statistic*
math ability at 7 relative to the lowest			
second quintile	0.069	0.056	1.23
third quintile	0.119	0.053	2.19
fourth quintile	0.142	0.053	2.62
highest quintile	0.139	0.054	2.55
reading ability at 7 relative to the lowest			
second quintile	0.087	0.055	1.55
third quintile	0.105	0.054	1.89
fourth quintile	0.161	0.055	2.86
highest quintile	0.217	0.055	3.77
missing ability at 7	0.289	0.072	3.44
math ability at 11 relative to the lowest			
second quintile	0.053	0.076	0.69
third quintile	0.102	0.077	1.31
fourth quintile	0.186	0.074	2.40
highest quintile	0.219	0.077	2.67
reading ability at 11 relative to the lowest			
second quintile	0.189	0.069	2.59
third quintile	0.200	0.071	2.66
fourth quintile	0.219	0.074	2.79
highest quintile	0.234	0.077	2.86
ability at 11 missing	0.344	0.066	4.39
white	-0.028	0.089	-0.31
number of siblings	-0.019	0.010	-1.88
father is a professional	0.220	0.063	3.17
went to Grammar school in 1974	0.153	0.050	2.94
went to Private school in 1974	0.149	0.058	2.45
mother's interest in education			
expects too much	0.007	0.079	0.09
is very interested	0.063	0.049	1.27
has some interest	-0.031	0.044	-0.72
father's interest in education			
expects too much	0.253	0.090	2.38
is very interested	0.031	0.042	0.74
has some interest	0.072	0.035	2.04
parental interest missing	0.017	0.081	0.21

We also control for school type, family background, and region. We report estimates for the effect of being white, the number of siblings, whether the father is professional and whether the child attended Grammar or Private school. $N = 1977$, Pseudo $R^2 = 0.136$.

Table 3: Probit estimates.

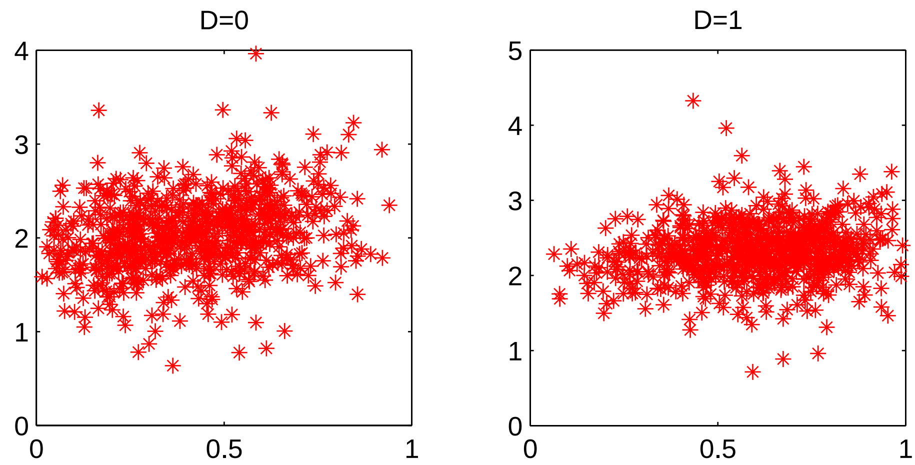

Figure 3: Scatter plot of log hourly wages against the propensity score.

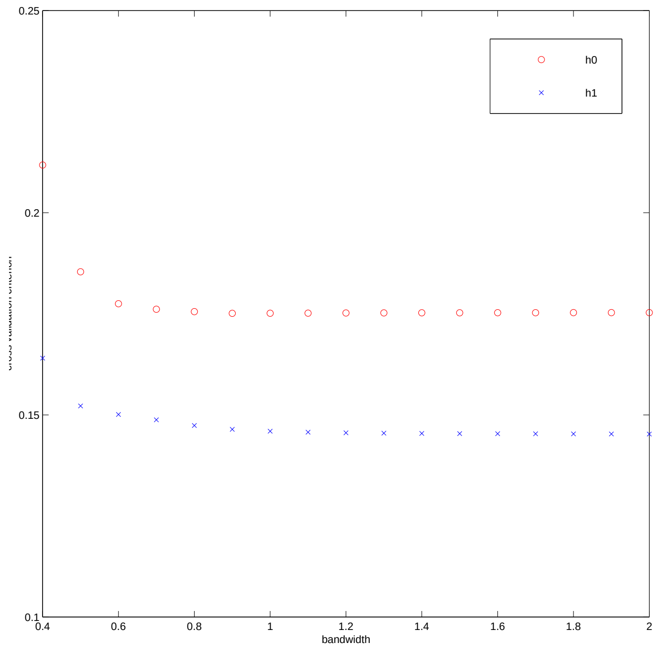

Figure 4: Cross validation crition. We let $h_0^{CV} = 0.9$ and $h_1^{CV} = 1.1$.

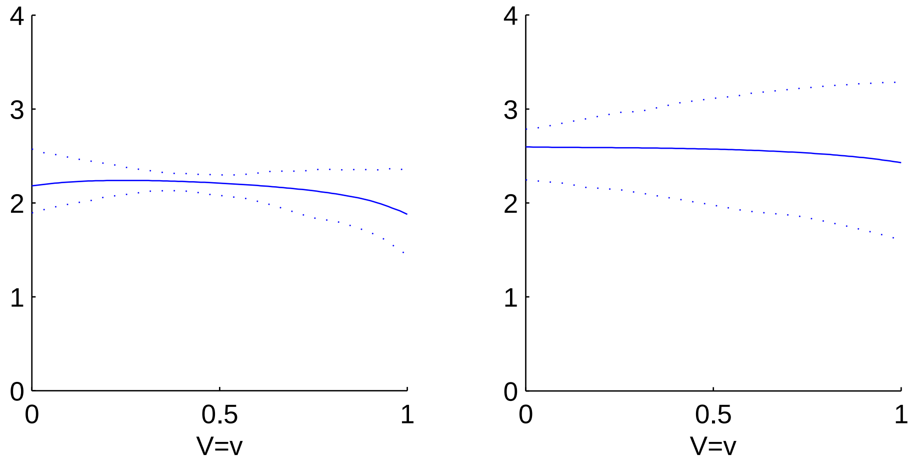

Figure 5: Level estimates for wages without (left) and with (right) a higher education degree against V. The dotted lines are bootstapped 95% confidence intervals.

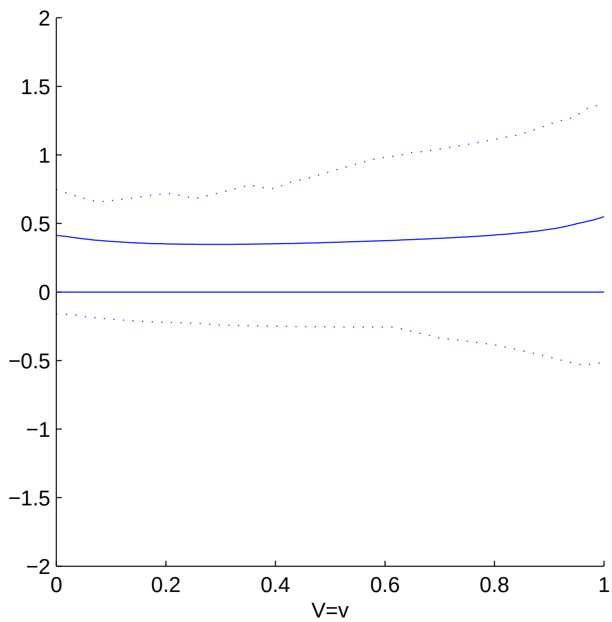

Figure 6: Marginal treatment effect and bootstapped 95% confidence interval gainst V.

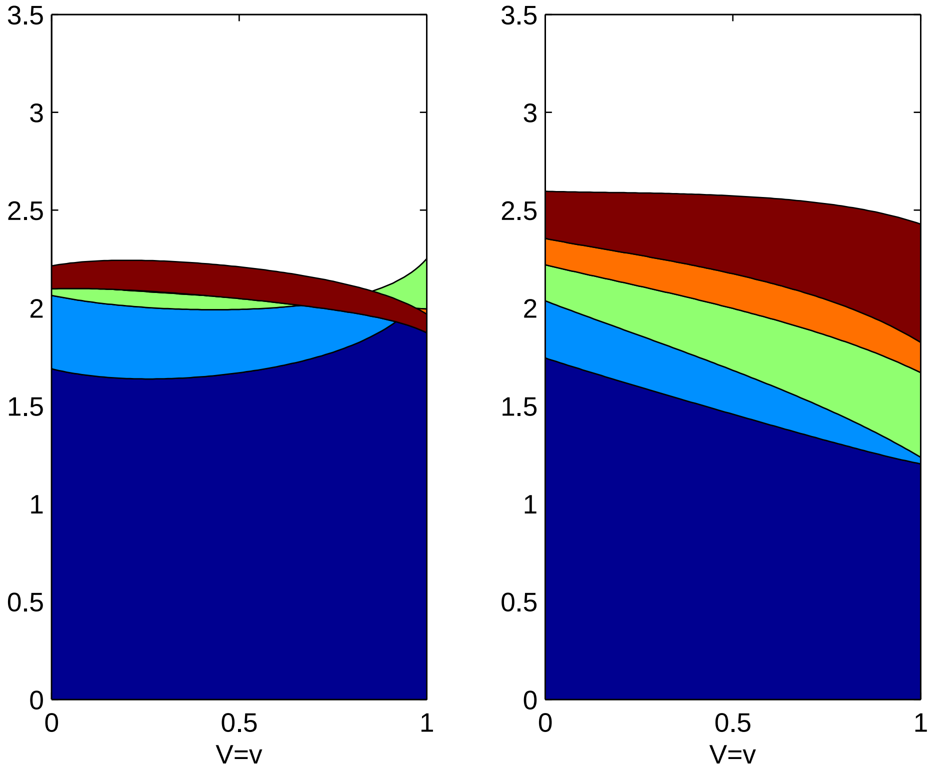

Figure 7: Decomposition of wages into *ceteris paribus* effects of constant and family background, London area rather than Wales, Comprehensive school rather than Private school, the father being intermediate, and measured math ability (from bottom to top) against V. Note that the green are in the top right corner of the left figure is a negative contribution, see also Figure 5.

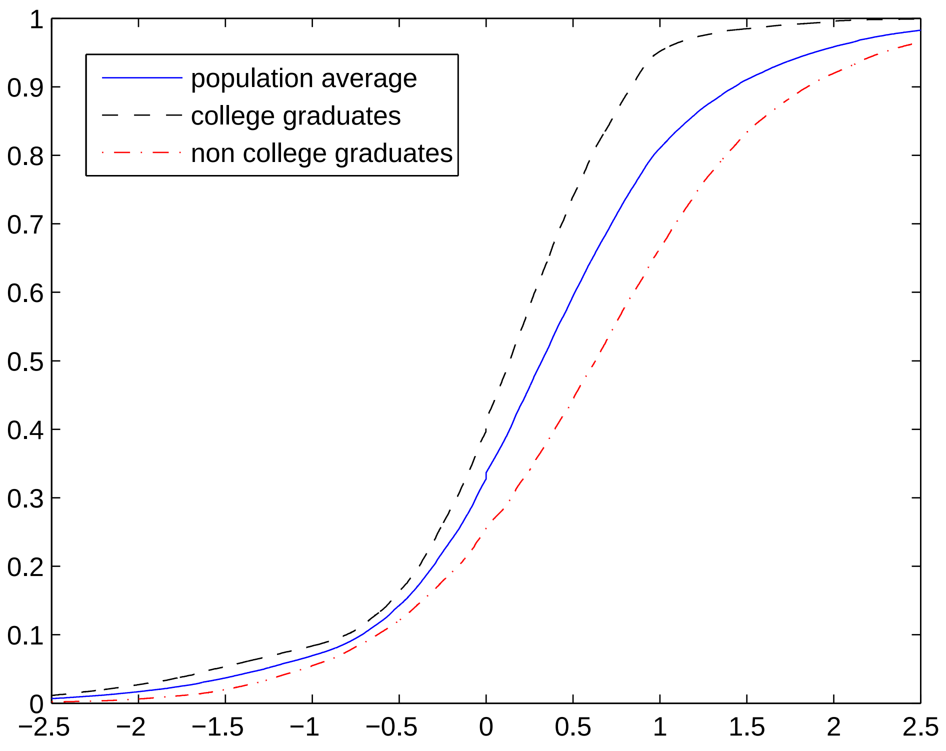

Figure 8: Cumulative distribution functions of simulated marginal treatment effects for the population, those who attended college (the treated), and those who did not (the untreated). Values of X were obtained from the observations and values of V were drawn (10 repetitions per observation).

References

Aakvik, A., J. J. Heckman, and E. J. Vytlacil (forthcoming): "Treatment Effects for Discrete Outcomes When Responses to Treatment Vary Among Observationally Identical Persons: An Application to Norwegian Vocational Rehabilitation Programs," *Journal of Econometrics*.

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.

Arrow, K. J. (1974): "Higer Education as a Filter," in *Efficiency in Universities*. Elsevier, New York.

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

Becker, G. S. (1964): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Columbia University Press, New York.

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

——— (1993): "Human Capital Revisited," in *Human Capital*. University of Chicago Press, 3rd edn., reprint of the 1989 Ryerson Lecture at the University of Chicago.

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 (forthcoming): "Estimating the Returns to Education: Models, Methods and Results," *Journal of the Royal Statistical Society*.

Blundell, R., and J. L. Powell (2001): "Endogeneity in nonparametric and semiparametric regression models," in *Advances in Econometrics, Proceedings of the Word Meetings, 2000*, ed. by L. Hansen, Amsterdam. North Holland.

——— (2003): "Endogeneity in nonparametric and semiparametric regression models," in *Advances in Econometrics, Proceedings of the Word Meetings, 2000*, 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 (2003): "Estimating the Return to Schooling When It Varies Among Individuals," Mimeograph.

Carneiro, P., and S. Lee (2004): "Comparative Advantage and Schooling," Mimeograph.

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.

——— (forthcoming): "Nonparametric Identification Under Discrete Variation," *Econometrica*.

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," Mimeograph.

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.

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.

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 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," Mimeograph, 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.

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.

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.

Solmon, L. C., and P. J. Taubman (1973): *Does College Matter?* Academic Press, New York.

Spence, M. (1972): "Market Signalling," Ph.D. thesis, Department of Economics, Havard University.

Stiglitz, J. E. (1973): "The Theory of 'Screening,' Education, and the Distribution of Income," Cowles Foundation Discussion Paper 354.

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.

——— (1973): "Higher Education, Mental Ability, and Screening," *Journal of Political Economy*, 8(1), 28–55.

——— (1974): *Does College Matter?* McGraw Hill, New York.

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.