# College Education and Wages in the U.K.:

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

Tobias J. Klein

University of Mannheim

ZEW Evaluation Conference
October 27, 2006

# Motivation

- How can we explain the differences in earnings between higher education graduates and a comparison group?
- Modelling unobservable factors is key in trying to understand these issues:
  - there might be confounding factors which influence both the decision whether to attend college or not, and earnings
  - these confounding factors are likely to be unobservable.

$\Rightarrow$ A correlated random coefficients model is semiparametrically estimated.

# Layout of the Presentation

- Econometric Model and Assumptions
- Theoretical Results: Identification and Estimation
- Application: The Determinants of Earnings in the U.K., in particular we focus on
  - the dependencies between wages,
  - unobserved ability,
  - social background,
  - and the decision whether to attend college.

# Econometric Model

## Model

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

$$\text{selection:} \quad D = \mathbb{1}\{P(Z) \geq V\} \tag{2}$$

$Y$ ... wage, $K$-vector $X = (1, X'_{-1})'$ constant and exogenous covariates $X_{-1}$, $D$ indicator for college, $Z$ instruments, $U$ unobservable "luck", and $V$ unobservable confounding factors

## Assumption 1 (Stochastic Restrictions)

*(i)* $(U, V)$ are jointly independent of $(X, Z)$ and *(ii)* $U$ is independent of $V$.

This is a correlated random coefficient model (see, e.g., Heckman and Vytlacil 1998). However, we allow for the dependence of $\varphi(D, U, V)$ on $V$.

# Parameters of Interest

- Among others, the expected level of earnings for a given $D$, $X$, and $V$,

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

- the marginal treatment effect

$$x'\mathbb{E}[\varphi(1, U, v) - \varphi(0, U, v)],$$

- and the average *ceteris paribus* effect of changes in $X_k$ for a given $D = d$, $X = x$, and $V = v$ answering, e.g., the question "How do wages relate to social background for a given level of unobserved ability?":

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

Note that the dependence of these parameters on $V$ is of economic importance.

# Related Results

## Nonparametric Identification (Heckman and Vytlacil)

Identification at $V = p$, where $p$ is a limit point of the support of $P$ conditional on $X$.

(We call $p$ a limit point of the support of $P$ if the density of $P$ is continuous in a neighborhood around $p$ and bounded away from zero.)

## Semiparametric Estimation of the Additive Model (Carneiro, Heckman and Vytlacil / Carneiro and Lee; CHV/CL thereafter)

$$
\begin{aligned}
Y &= \mu(D, U, V) + X'_{-1}\gamma(D, U) \\
D &= \mathbb{1}\{P(Z) \geq V\}
\end{aligned}
$$

with $(X, Z) \perp\!\!\!\perp (U, V)$ and some technical conditions hold. Identified at $V = p$, where $p$ is a limit point of the support of $P$.

# Contribution of this Paper

Consider the wage equation in CHV/CL and the one that is proposed:

$$Y = \mu(D,U,V) + X'_{-1}\gamma(D,U) \quad (3)$$

$$Y = \varphi_1(D,U,V) + X'_{-1}\varphi_{-1}(D,U,V). \quad (4)$$

$(3)$ is more general than $(4)$ because we can always set $\varphi_1(D,U,V) = \mu(D,U,V)$ and $\varphi_{-1}(D,U,V) = \gamma(D,U)$.

$\Rightarrow$ We do not restrict the effect of $X$ on $Y$ to be independent of $V$. This generalizes the approach to estimation taken by CHV/CL.

The model implies that

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

**Theorem 1 (similar to Carneiro and Lee)**

Under Assumption 1 and 2 and some regularity conditions the conditional average structural function is identified at $V = p$, where $p$ is a limit point of the support of $P$, and given by

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

# Estimation

We have established that

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

This is a Varying-coefficient Model (Cleveland, Grosse and Shyu 1991, Hastie and Tibshirani 1993). We estimate the coefficient function by local linear smoothing (Fan and Zhang 1999, Xia and Li 1999 in a time series context, Christopeit and Hoderlein 2006). From these estimates we calculate estimates of the conditional average structural function using the formulas in Theorem 1.

# Application: Data

- National Child Development Survey (NCDS), British Cohort Data: detailed records for all those who were born between 3rd and 9th of March, 1958; Waves in 1965, 1969, 1974, 1981 and 1999/2000.
- Outcome of interest: log hourly wages in 1981 (age of 33).
- We select males who at least completed their A-levels or a similar degree. $51.4\%$ of them are higher education graduates, $N = 1501$.
- Following Blundell, Dearden, and Sianesi (2005, BDS in the remainder) we include the mother's and father's interest in the education of the child in $Z$ additional to all variables in $X$.

# Average Returns to College Education

| | 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: 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).

# *Ceteris Paribus* Effects 1/2

| | average *ceteris paribus* effect | unobserved heterogeneity | bias from imposing additivity |
|---|---|---|---|
| | | NO COLLEGE DEGREE | |
| father professional | 0.293 | yes | -0.316 |
| father's years of edu. | 0.034 | no | no |
| | | COLLEGE DEGREE | |
| father professional | 0.794 | yes | -0.316 |

Table: Effects that are significant at the 95 per cent level.

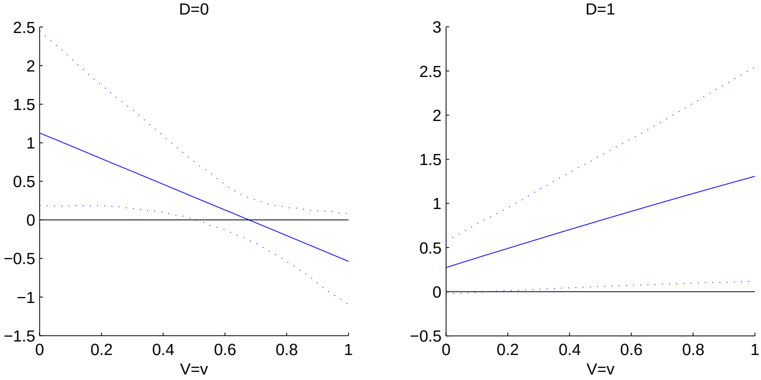


Figure: Conditional average *ceteris paribus* effect of the father being professional. Point estimates and bootstrapped 95% confidence intervals.

# Sorting 1/4

- Sorting based on comparative advantage if the marginal treatment effect is falling in $V$.
- If wage levels are increasing in $V$, we have

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

# Sorting 2/4

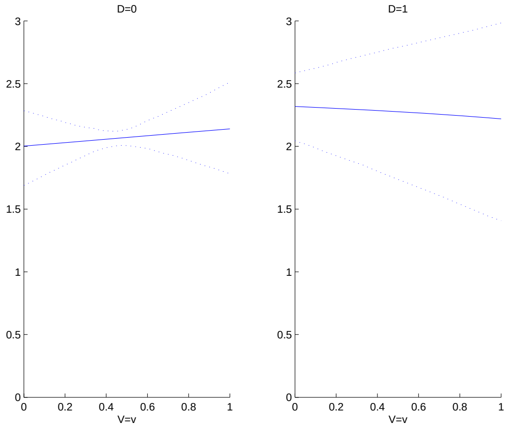


Figure: Point estimates and bootstrapped 95% confidence intervals of the conditional average structural function. Reported for a representative individual with median characteristics.

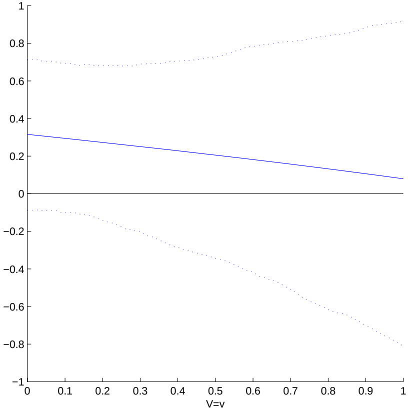


Figure: Point estimates and bootstrapped 95% confidence intervals of the marginal treatment effect. Reported for a representative individual with median characteristics.

# Sorting 4/4

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

# Summary and Conclusions 1/2

- We have proposed and implemented a semiparametric estimator for expected wage levels and their dependence on the endogenous schooling choice.
- Virtue: Dimensionality reduction along the dimension of the observables while not imposing any limiting restrictions on the joint distribution of unobservables.
- Allows for nonseparabilities.

# Summary and Conclusions 2/2

- Empirical Results:
  - Monetary returns to a college are sizable (the population average effect is estimated to be 0.46).
  - Returns are lowest for those who attend college and highest for those who do not. Sorting based on comparative advantage with respect to monetary returns is not supported.
  - Differences in wages can be traced back to observables (measured math ability), unobservables (unobserved ability), and the combination of the two.
  - The effect of observables (social class of the father, e.g.) depends on unobservables.

# Estimator

Xia and Li (1999): Assume that the coefficient functions $\beta_k(d,p)$, $k = 1, \ldots, K$, are bounded and have bounded second derivatives. By a Taylor expansion,

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

where $\bar{p}$ is a point between $p$ and $\tilde{p}$.

Select all observations with $D = d$. Index observations 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.

# Details for Theorem 1

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).

**Assumption 2 (Regularity Conditions)**

(i) All first moments exist and (ii) the distribution of $V$ is absolutely continuous with respect to Lebesgue measure.

(Note that we call $p$ a limit point of the support of $P$ if the density of $P$ is continuous in a neighborhood around $p$ and bounded away from zero.)

# Proof of Theorem 1

This proof is similar to the one of Carneiro and Lee (2004). 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.
Normalize $V$ 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.