NBER WORKING PAPER SERIES

SELECTION BIAS, COMPARATIVE ADVANTAGE AND HETEROGENEOUS RETURNS TO EDUCATION: EVIDENCE FROM CHINA IN 2000

James J. Heckman
Xuesong Li

Working Paper 9877
http://www.nber.org/papers/w9877

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2003

James J. Heckman is Henry Schultz Distinguished Service Professor of Economics at the University of Chicago and a Senior Fellow of the American Bar Foundation, E-Mail: jjh@uchicago.edu. Xuesong Li is a Research Fellow at the Institute of Quantitative & Technical Economics (IQTE) of the Chinese Academy of Social Sciences (CASS), he is currently a visiting scholar at the Department of Economics of the University of Chicago, E-Mail: lixs@public.bta.net.cn. We thank Salvador Navarro-Lozano for his help with the program used in this work. We thank Leslie Cooke for research assistance. We thank Pedro Carneiro, Edward Vytlacil and Junsen Zhang for their helpful comments. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research

Selection Bias, Comparative Advantage and Heterogeneous Returns to Education:
Evidence from China in 2000
James J. Heckman and Xuesong Li
NBER Working Paper No. 9877
July 2003
JEL No. C31

ABSTRACT

This paper uses newly available Chinese micro data to estimate the return to college education for late 20th century China when allowing for heterogeneous returns among individuals selecting into schooling based on these differences. We use recently developed semiparametric methods to identify the parameters of interest. We demonstrate that heterogeneity among people in returns to schooling is substantial. People sort into schooling on the basis of the principle of comparative advantage, which we document to be an empirically important phenomenon in modern Chinese labor markets. Standard least squares or instrumental variable methods do not properly account for this sorting. Using new methods that do, we estimate the effect on earnings of sending a randomly selected person to college is a 43% increase in lifetime earnings (nearly 11% annually) in 2000 for young people in urban areas of six provinces of China. The effect of college on those who go is 13%. Our evidence, and simple least squares evidence, suggests that after 20-plus years of economic reform with market orientation, the return to education has increased substantially in China, compared to the returns measured in the 1980's and the early 1990's.

James J. Heckman
University of Chicago
Department of Economics
1126 East 59th Street
Chicago, IL 60637
and NBER
jjh@uchicago.edu

Xuesong Li
Institute of Quantitative & Technical Economics
Chinese Academy of Social Sciences (CASS)
No. 5 Jianguomennei Street
Beijing 100732 P. R. China
lixs@public.bta.net.cn

1. Introduction

Heterogeneity and missing counterfactual states are central features of microdata. Due to unobserved heterogeneity, observationally identical people make different choices, earn different wages and hold different levels and compositions of asset portfolios. The evaluation problem for social programs arises from a missing data problem. We cannot observe the outcomes of all possible choices for the same person. If we observe wages for college graduates, we cannot observe the wages they would have earned if they had been high school graduates.

Conventional approaches to selection and missing data problems do not account for heterogeneity in responses to schooling on which agents select into schooling. This paper uses newly released cross-sectional micro data from the China Urban Household Investment and Expenditure Survey (CUHIES 2000), to estimate the return to education for China when responses to schooling differ among individuals and individuals select into schooling based their idiosyncratic returns. Our work draws on previous research by Heckman and Vytlacil (1999, 2000, 2001), Carneiro, Heckman and Vytlacil (2001) and Carneiro (2002), which develops a semiparametric framework that accounts for heterogeneity and selection.[1]

Our results reveal that the average treatment effect (*ATE*) of four year college attendance (the earnings gain arising from randomly selecting someone to go to college for four years), is 43% (the annual return is 10.8%) in 2000 for young people in urban areas of six provinces of China, whereas the *OLS* (Ordinary Least Squares) and *IV* (Instrumental Variables) estimators give 29% and 56% respectively (with estimated annual returns of 7.25% and 14% respectively). Heterogeneity in returns is substantial in the population. Estimated selection bias is an empirically important negative 22%. Like Carneiro, Heckman and Vytlacil (2001), we find that there is comparative advantage in the labor market for schooling. The best college graduates are among the worst high school graduates. *OLS* gives a downward-biased estimate of *ATE*. *IV* produces an upward biased estimate of *ATE*.

[1] The *MTE* is the central concept in this literature. It was introduced by Bjorklund and Moffitt (1987). The marginal treatment effect is the average return to schooling for persons indifferent to going on to schooling at different levels of unobservable factors that determine schooling choices. Heckman and Vytlacil (1999, 2000, 2001) show that all conventional treatment parameters are different weighted averages of this parameter.

After more than twenty years of economic reform with market orientation, the average return to education in China measured by *OLS* or *ATE* has increased markedly when compared to those in the 1980's and early 1990's. (Chow 2001 presents estimates of *OLS*-generated rates of return in this period). Education markets have begun to function effectively in China, and skills are now being rewarded more adequately than they have been in the past.

The paper is organized in the following way. Section 2 describes earnings models with and without heterogeneous returns to education. Section 3 defines selection bias, defines the marginal treatment effect and presents a semiparametric method for estimating it. Section 4 discusses our data and presents empirical results for China. Section 5 concludes.

2. Models with and without Heterogeneity

We first consider a conventional model of the return to education without heterogeneity in returns. We write the following common coefficient Mincer model:

$$\ln Y_i = \beta S_i + \gamma X_i + U_i \tag{1}$$

where i is a subscript for individuals ($i = 1, 2, \cdots, \mathrm{n}$), $\ln Y_i$ is log income, S_i is schooling level or years of schooling, X_i is a vector of variables such as an intercept, years of Mincer experience, Mincer experience squared, and dummy variables for sex, region, sector, and ownership of firm. U_i is the residual term with $E(U_i) = 0$, β is the rate of return to education, and γ is a vector of coefficients.

One problem with *OLS* (Ordinary Least Squares) estimates of Equation (1) is that there may be an omitted ability A_i, which is in the residual term U_i. Many empirical analysts suspect that $Cov(A_i, S_i) \neq 0$ so that $E(U_i | S_i) \neq 0$ and *OLS* gives biased and inconsistent estimates (Griliches, 1977 is a classic statement of this problem).[2]

[2] Most data sets do not contain measures of ability. Economists use three strategies to eliminate or attenuate the ability bias. A huge literature uses instrumental variables (*IV*). The goal of this literature is to find an instrument I_i that is highly correlated with S_i but not correlated with U_i. The second approach uses the fixed effect method: find a paired comparison

The specification we consider is more general than the model (1). We estimate a model with heterogeneous returns to education, which may be written in random coefficient form as

$$\ln Y_i = \beta_i S_i + \gamma X_i + U_i \tag{2}$$

where β_i is the heterogeneous rate of return to education, which varies among individuals. X_i is a vector of conditioning variables defined below. This model accounts for ability bias in a more general setting.

In this paper we focus on two schooling choices: high school and college. We let $S_i = 1$ denote four-year college graduates and $S_i = 0$ for senior high school graduates (those not going to college). Clearly, there are more choices of schooling and our analysis is a simplification of reality, but is a natural starting point with ample precedents in the literature. There is considerable evidence in many contexts that returns to schooling are nonlinear in years of schooling so conventional log wage on years of schooling regression coefficients generate rates of return that are badly biased estimates of the return to college education. (Heckman, Lochner and Todd, 2003).

The two potential selection outcomes $(\ln Y_{0i}, \ln Y_{1i})$ can be written as

$$\begin{cases} \ln Y_{0i} = \gamma_0 X_i + U_{0i} \quad \text{if} \quad S_i = 0 & (3a) \\ \ln Y_{1i} = \gamma_1 X_i + U_{1i} \quad \text{if} \quad S_i = 1 & (3b) \end{cases}$$

where $E(U_{0i} | X_i) = 0$ and $E(U_{1i} | X_i) = 0$ in the population. In the notation of equation (2), observed log earnings $Y_i = S_i \ln Y_i + (1 - S_i) \ln Y_{0i}$ and $U_i = S_i U_{1i} + (1 - S_i) U_{0i}$.

---

such as a genetic twin or sibling with similar or identical ability. A third approach is to use proxy variables for ability and include them as regressors in X_i.

Many data sets do not have enough information to use the fixed effect method, and the method is critically dependent on additive separability of errors. Such comparisons may exacerbate measurement error problems. It is also very hard to find satisfactory instruments. In fact, most commonly used instruments in the schooling literature are invalid because they are correlated with the omitted ability. (See Carneiro and Heckman, 2002 and Carneiro, 2002).

An alternative method uses proxies for ability and includes them as the regressors. Many empirical analyses reveal that better family background and better family resources are usually associated with better environments that raise ability (Carneiro and Heckman, 2003). We use parental income as a proxy for ability in our empirical work.

In a cross section it is usually impossible to know both $\ln Y_{0i}$ and $\ln Y_{1i}$ for anyone due to a fundamental missing data problem. For those going to college, we cannot observe $\ln Y_{0i}$; for those not going to college, we cannot observe $\ln Y_{1i}$. So we can only determine the distributions $F(\ln Y_{0i} \mid X_i, S_i = 0)$ and $F(\ln Y_{1i} \mid X_i, S_i = 1)$ but not $F(\ln Y_{0i} \mid X_i)$ or $F(\ln Y_{1i} \mid X_i)$. In the presence of heterogeneity and selection in general, we can no longer use conventional methods like *OLS* or Instrumental Variables (*IV*) to identify economically interesting parameters.

Collecting results

$$\ln Y_i = S_i \ln Y_{1i} + (1 - S_i) \ln Y_{0i} \tag{4a}$$

$$= [(\gamma_1 - \gamma_0) X_i] S_i + \gamma_0 X_i + [U_{0i} + (U_{1i} - U_{0i}) S_i] \tag{4b}$$

$$= [(\gamma_1 - \gamma_0) X_i + (U_{1i} - U_{0i})] S_i + \gamma_0 X_i + U_{0i} \tag{4c}$$

$$= \beta_i S_i + \gamma_0 X_i + U_{0i} \tag{4d}$$

where

$$\beta_i = (\gamma_1 - \gamma_0) X_i + (U_{1i} - U_{0i}) \tag{5}$$

is the heterogeneous return to education for individual i. When $\gamma_1 \neq \gamma_0$ (i.e. there is an observed heterogeneity term $(\gamma_1 - \gamma_0) X_i$), or $U_{1i} \neq U_{0i}$ (i.e. there is an unobserved heterogeneity term $(U_{1i} - U_{0i})$), β_i varies in the population, the return to schooling is a random variable with a distribution. In the first case where we condition on *X*, the distribution of returns is degenerate. In the second case it is not degenerate. The mean of β_i given *X* is:

$$\overline{\beta}(X) = E(\beta_i \mid X_i) = E[(\gamma_1 - \gamma_0) X_i] \tag{6}$$

Suppose individuals select going to college (or not) according to the following decision rule:

$$\begin{aligned} &S_i^* = P_i(Z_i) - U_{si} \\ &S_i = 1 \quad \text{if} \quad S_i^* \geq 0 \\ &\quad\;\; = 0 \qquad \text{otherwise,} \end{aligned} \tag{7}$$

where S_i^* is a latent variable denoting the net benefit of going to school and Z_i is an observed vector of variables (Z_i may include some X_i). $P_i = P_i(Z_i)$ is the propensity score or probability of receiving treatment (going to college), which can be estimated by a logit or probit model. U_{si} is the unobserved heterogeneity for individual i in the treatment selection equation.

Without loss of generality we may assume that $U_{si} \sim Unif[0,1]$ (See Heckman and Vytlacil, 1999). The decision of whether to go to college (or not) for individual i is determined completely by the comparison of the observed heterogeneity $P_i(Z_i)$ with the unobserved heterogeneity U_{si}. The smaller the U_{si}, the more likely the person goes to college.

3. Selection Bias and The Marginal Treatment Effect

Let $\Delta_i = \ln Y_{1i} - \ln Y_{0i}$ be the economic (gross) return to a policy that moves individual i from $S_i = 0$ to $S_i = 1$. According to Equations (3a), (3b) and (5), $\Delta_i = \beta_i$, is the causal effect of education. Using equations (3a), (3b) and (6), the probability limit of the ordinary least squares estimator can be written as:

$$\begin{aligned}\text{plim}(\hat{\beta}_{OLS}) &= E(\ln Y_i | X_i,\ S_i = 1) - E(\ln Y_i | X_i,\ S_i = 0)\\ &= E(\gamma_1 X_i + U_{1i} | X_i,\ S_i = 1) - E(\gamma_0 X_i + U_{0i} | X_i,\ S_i = 0)\\ &= \underset{\textit{(ATE)}}{\bar{\beta}(X)} \quad + \quad \underset{\textit{(Bias)}}{[E(U_{1i} | S_i = 1) - E(U_{0i} | S_i = 0)]}\end{aligned} \tag{8}$$

where ATE is the average treatment effect, (the effect of randomly assigning a person with characteristics X to schooling) defined as

$$ATE = E(\Delta_i | X_i) = E(\beta_i | X_i) = \bar{\beta}(X)\,. \tag{9}$$

If agents know and act on some components of (U_{0i}, U_{1i}), S_i is generally correlated with both U_{0i} and U_{1i}, and the second term in Equation (8) will be not zero, so OLS is biased for ATE.

Note that Equation (8) can also be written as:

$$\begin{aligned}\text{plim}(\hat{\beta}_{OLS}) &= E(\ln Y_i | X_i,\ S_i = 1) - E(\ln Y_i | X_i,\ S_i = 0)\\ &= \underset{\textit{(TT)}}{E(\beta_i | X_i,\ S_i = 1)} \quad + \quad \underset{\textit{(Selection Bias)}}{[E(U_{0i} | S_i = 1) - E(U_{0i} | S_i = 0)]}\end{aligned} \tag{10}$$

where TT is treatment on the treated, the effect of treatment on those who receive it (*e.g.*, goes to college) compared to what they would experience without treatment (*i.e.*, do not go to college), defined as:

$$TT = E(\Delta_i | X_i,\ S_i = 1) = E(\beta_i | X_i,\ S_i = 1)$$
$$= \overline{\beta}(X) + E(U_{1i} - U_{0i} | S_i = 1) = ATE + E(U_{1i} - U_{0i} | S_i = 1) \qquad (11)$$
$$(\textit{Sorting Effect})$$

The sorting effect $E(U_{1i} - U_{0i} | S_i = 1)$ is the mean gain of the unobservables for people who choose "1". The selection bias $E(U_{0i} | S_i = 1)$ - $E(U_{0i} | S_i = 0)$ is the mean difference in the no schooling (S=0) unobservables between those who go to school and those who do not. It is the difference in unobservables between what college graduates would earn if they were high school graduates and what high school graduates would earn. The bias in (8) is the sum of sorting and selection bias.

Conventional IV estimators do not, in general, identify these treatment parameters in the presence of heterogeneity and selection. Finding an instrument I_i correlated with S_i but not U_{0i} or even $U_{1i} - U_{0i}$ is not enough to identify $\overline{\beta}(X)$, because:

$$\text{plim}\hat{\beta}_{IV} = \frac{Cov(I_i, \ln Y_i)}{Cov(I_i, S_i)} = \overline{\beta}(X) + \frac{Cov(I_i, U_{0i})}{Cov(I_i, S_i)} + \frac{Cov[I_i, (U_{1i} - U_{0i})S_i]}{Cov(I_i, S_i)}$$
$$= \overline{\beta}(X) + \frac{Cov[I_i, (U_{1i} - U_{0i})S_i]}{Cov(I_i, S_i)} = \overline{\beta}(X) + \frac{Cov[I_i, (U_{1i} - U_{0i}) | S_i = 1]P_i}{Cov(I_i, S_i)} \qquad (12)$$

where $P_i = \Pr(S_i = 1)$ is the propensity score. In the presence of both heterogeneity and selection bias, $U_{1i} \neq U_{0i}$, $U_{1i} - U_{0i}$ is dependent on S_i, so the second term in Equation (12) will be not zero, thus $\text{plim}(\hat{\beta}_{IV}) \neq \overline{\beta}(X)$ so IV is not a consistent estimator. Only in some very special circumstances, when $U_{1i} - U_{0i} = 0$ (i.e. neither unobserved heterogeneity nor selection bias exist) or when $U_{1i} \neq U_{0i}$ but $U_{1i} - U_{0i}$ is independent of S_i (i.e. there is unobserved heterogeneity but no selection bias), will the second term in Equation (12) be zero. In this case, IV is a consistent estimator for $\overline{\beta}(X)$ (Heckman, 1997 and Heckman and Navarro-Lozano, 2003).

Neither OLS nor IV is a consistent estimator of the mean return to education in the presence of heterogeneity and selection. However, under the assumptions presented in Heckman and Vytlacil (1999, 2000, 2001), Carneiro, Heckman and Vytlacil (2001), Carneiro (2002) and Navarro-Lozano (2002), it is possible to identify the heterogeneous return to education with marginal treatment effect (MTE) via the method of Local Instrument Variables (LIV), where MTE is:

$$
\begin{aligned}
MTE(X_i = x, U_{si} = u_s) = E(\Delta_i \left| X_i = x, U_{si} = u_s\right.) = E(\beta_i \left| X_i = x, U_{si} = u_s\right.) \\
= (\gamma_1 - \gamma_0)x + E(U_{1i} - U_{0i} \left| U_{si} = u_s\right.)\ .
\end{aligned}
\tag{13}
$$

The MTE is the average willingness to pay (WTP) for $\ln Y_{1i}$ (compared to $\ln Y_{0i}$) given characteristics X_i and unobserved heterogeneity U_{si}.[3] MTE can be estimated from the following relationship, where LIV can be estimated by semiparametric methods for derivatives (see Heckman 2001):

$$
MTE(X_i = x, U_{si} = P_i = p) = LIV(X_i = x,\ P_i = p) = \frac{\partial E(\ln Y_i \left| X_i = x,\ P_i = p\right.)}{\partial p}\ . \tag{14}
$$

For notational simplicity, we keep the conditioning on X implicit in what follows.

Heckman and Vytlacil (1999, 2000, 2001) and Carneiro (2002) establish that all the other treatment variables can be unified using MTE

$$ATE = \int_0^1 MTE(u_s)du_s \qquad \text{(Average Treatment Effect)}$$

$$TT = \int_0^1 MTE(u_s)h_{TT}(u_s)du_s \qquad \text{(Treatment on the Treated)}$$

$$TUT = \int_0^1 MTE(u_s)h_{TUT}(u_s)du_s \qquad \text{(Treatment on the Untreated)}$$

Where the weights are:

$$h_{ATE}(u_s) = 1$$

$$h_{TT}(u_s) = \frac{1 - F_P(u_s)}{E(P_i)} = \frac{\int_{u_s}^1 f(p)dp}{E(P_i)}$$

$$h_{TUT}(u_s) = \frac{F_P(u_s)}{E(1 - P_i)} = \frac{\int_0^{u_s} f(p)dp}{E(1 - P_i)}\ .$$

Treatment on the untreated (TUT) is the effect of treatment on those who do not receive it (i.e. do not go to college) compared to what they would experience with the treatment (i.e. go to college), which is defined as:

$$
\begin{aligned}
TUT = E(\Delta_i \left| X_i,\ S_i = 0\right.) = E(\beta_i \left| X_i,\ S_i = 0\right.) \\
= \overline{\beta} + E(U_{1i} - U_{0i} \left| S_i = 0\right.).
\end{aligned}
\tag{16}
$$

[3] MTE was introduced into the literature in a parametric context by Anders Bjorklund and Robert Moffitt (1987).

4. Data Set and Empirical Results

Our data are from the China Urban Household Income and Expenditure Survey (CUHIES) for the year 2000, which was conducted by the Urban Socio-Economic Survey Organization of the National Bureau of Statistics. The survey is a sequence of cross-sections from 1992 to 2002 and is ongoing. The urban data randomly selects households across the whole urban population.

We have the data for the year 2000 for urban areas of six provinces: Guangdong, Liaoning, Sichuan, Shaanxi, Zhejiang and Beijing. Four provinces, Guangdong, Zhejiang, Beijing and Liaoning, are located in eastern part of China, while the other two, Sichuan and Shaanxi, are in the western part. Table 1 provides the comparison of average resident income in urban areas among these provinces. The average resident incomes in the three provinces, Beijing, Guangdong and Zhejiang, are much higher than the average level of China, while they are a bit lower in the other three provinces, Sichuan, Liaoning and Shaanxi. The average income in the six provinces we use is 7627 yuan, which is higher than the average income of China, 6280 yuan.

The sample size for the six provinces is 4250 households. For each household, there is rich information on all household members, including head, spouse, children and parents. Age, sex, education level, employment status and enterprise ownership, occupation, years of work experience and total annual income are available for each household member. There are seven education levels in the sample: university, college, special technical school, senior high school, junior high school, primary school, and other.

For our purposes, we combine all the children in the six provinces who are either college or university graduates or senior high school graduates. They are all working and earn positive wages in 2000. Our sample consists of 587 individuals, including 273 people with four-year college (or university) certificates and 314 people with only senior high school certificates. There are 331 males and 256 females in the sample. The summary statistics for the variables used in our analysis are reported in Table 2, which reveals the individuals in the sample are mainly young adults with a mean age of 26.3. Thus ours is an analysis of wages early in the life cycle of new cohorts of Chinese workers.

Table 3 presents OLS and IV estimates of the mean return to four-year college attendance. We use the probability of going to college as the instrument with the exclusions defined below. The OLS and IV estimates are 29% and 56% respectively for the young people in the urban areas of six provinces of China in 2000 (annualized 7.25% and 14%, respectively). The *OLS* estimates are much higher than the *OLS* estimates reported by Chow (2001) for an earlier period (1980's and early 1990's). The variables in the outcome equation include the years of Mincer experience, Mincer experience squared, our proxy of ability (we use parental income as the proxy in this paper) and some dummy variables such as the sex, the provinces of residence, the sector and the firm ownership in which he or she works. The propensity score is estimated by a logistic model, with coefficient estimates presented in Table 4.[4]

We use father's education, mother's education, parental income, and the year of birth as determinants of the probability of going to college. The last column of the table is the mean marginal effect for each explanatory variable. Figure 1 shows the density function for the estimated probability of college attendance $(\Pr(S=1))$.

Table 5 and Figure 2 give the results from our semiparametric estimation. We use parental income in the earnings function to control for ability. For details on the procedures used to generate these numbers see Carneiro, Heckman and Vytlacil (2001), whom we follow. Table 5 contains the estimated coefficients for Equations (3a) and (3b) using local linear regression. Figure 2 plots the estimated marginal treatment effect as a function unobserved heterogeneity u_s, the components in the choice equation. The MTE is declining in u_s. This implies that people with lower u_s, i.e. those more likely to go to college according to the decision rule (7), have higher marginal returns to schooling. The people with the highest u_s, who are least likely going to college, have the lowest average returns. Figure 2 suggests substantial heterogeneity in the return to education for China. The declining *MTE* implies that matching and conventional *OLS* and *IV* methods do not identify any relevant treatment effect in our data. (See Heckman, 2001 or Heckman and Navarro, 2003). It also suggests that the marginal participant in Chinese higher education earns less than the average participant.

Table 6 presents a comparison among various treatment parameters. The average return to 4-year college attendance for a randomly selected person is 43% (11% annually) given by ATE. The effect of going to college

[4] The general forms of the logistic model and the marginal effects derived from it are defined as:

$$\Pr(S=1 \mid X) = \frac{\exp(\beta' X)}{1+\exp(\beta' X)} = \Lambda(\beta' X)$$

$$\text{Marginal effects} = \frac{\partial \Pr(S=1 \mid X)}{\partial X} = \Lambda(\beta' X)[1-\Lambda(\beta' X)]\beta .$$

on those who go is 51%, so there is purposive sorting into schooling on the basis of gain (13% annual). The *OLS* estimator is downward biased for *ATE* with only a 29% return (7% annual). The inconsistent *IV* estimator is 56% and is upward biased due to heterogeneity and selection bias (Heckman and Vytlacil derive the exact bias). The Chinese data set show that $IV > ATE > OLS$. The estimated selection bias of –22% is very important in estimating the economic return to schooling for China. Persons who go to college would make poor high school graduates. Treatment on the treated (TT) and treatment on the untreated (TUT) are 51% and 36% respectively. Thus *IV* is upward biased for *TT*.

The estimated sorting gain is large and positive, suggesting that the principle of comparative advantage is important. The "sorting gain" reported in Table 6 is defined as:

$$\text{Sorting Gain } = E(\beta_i - \overline{\beta} \mid X_i, S_i = 1) = E(\beta_i \mid X_i, S_i = 1) - \overline{\beta}(X) = TT - ATE \tag{17}$$

Table 6 and Figure 2 also reveal that the average return to college attendance is high in 2000 for young people in urban areas of the six provinces of China.

Figure 3 plots the estimated weights used to form treatment parameters ATE, TT and TUT. ATE weights MTE evenly. TT overweights the ATE for persons with low values of u_s who, *ceteris paribus*, are more likely to attend college. TUT overweights the ATE for persons with high values of u_s who are less likely to attend college. Not surprisingly, in light of the shape of MTE and the shape of the weights, $TT > ATE > TUT$. This is also revealed in the Table 6. There is substantial heterogeneity among individuals and there is a positive sorting gain and a negative selection bias.

In order to test the importance of introducing a proxy for ability in the wage equation, we exclude parental income from the wage equation and re-estimate the marginal treatment effects. The results are displayed in Figure 4. In this case, the MTE increases in u_s and its average value is obviously much higher than that in Figure 2. Therefore, neglecting ability (or its proxy) results in an upward bias for the marginal treatment effect and the estimated return to schooling.

To explore the sensitivity of the estimates to various exclusions and inclusions, we present the estimates shown in Tables 7(a) – 7(d). *MTE*s are plotted for various specifications of the model controlling (or not controlling) for sectoral choices and for ability (see Figures 5 and 6).

Our main specification conditions on sectoral choices including the ownership structure of the firm. As is well known, conditioning on sectoral choices in the wage equation is likely to lead to an understatement of the full return to schooling because one benefit of education is that it facilitates choice of sector. When we drop various firm ownership and sectoral indicator variables, estimated returns go up (see Table 7d for estimates deleting all sectoral choice and ownership variables). This is clear from Figure 5. However, the effect of including or excluding these variables is very small on estimated marginal treatment effects. Failing to condition on ability (parental income) raises the estimated return to implausible levels and changes the shape of the estimated *MTE*. This evidence is consistent with the findings of Carneiro (2002).

5. Concluding Remarks

This paper uses newly available micro data to identify the returns to higher education in China. We demonstrate the importance of considering heterogeneity and selection bias. Neglecting these two factors leads to biased and inconsistent estimates such as those obtained using conventional *OLS* and *IV* parameters. We demonstrate the importance of proxying for ability in the wage equation to identify returns to education. Excluding it leads to implausibly high estimates of the return to schooling. On the other hand, controlling for sectoral choices barely budges the estimates.

In 2000, the average return to four-year college attendance is 43% (on average 11% annually) for young people in the urban areas of the six provinces. The returns to those going to college are even higher. These estimates are all higher than the conventional *OLS* estimates of the Mincer model, which in turn are higher than the *OLS* estimates reported for earlier time periods. They imply that, after 20-plus years of economic reform with market orientation, the average return to education in China has increased substantially when compared to those in the 1980's and early 1990's.

References

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

Carneiro, P. (2002), "Heterogeneity and Selection in the Return to Schooling: Implications for Education Policy Evaluation", Ph.D. Dissertation, University of Chicago.

Carneiro, P. and J. Heckman (2002), "The Evidence on Credit Constraints in Post-Secondary Schooling", *Economic Journal*, Volume 112, issue 482, pp:705-734.

_____. (2003), "Human Capital Policy," forthcoming in J. Heckman, A. Krueger, *Inequality in America: What Role for Human Capital Policies*, Cambridge, MA: MIT Press.

Carneiro, P., J. Heckman and E. Vytlacil (2001), "Estimating the Return to Education When It Varies Among Individuals", Working paper, University of Chicago.

China's Urban Household Income and Expenditure Survey (CUHIES) (2000), National Bureau of Statistics, Urban Socio-Economic Survey Organization.

Chow, G. (2001), *China's Economic Transformation*. Malden, MA: Blackwell Publishers.

Griliches, Z. (1977), "Estimating the Returns to Schooling: Some Econometric Problems," Econometrica, Volume 45, number 1, pp. 1-22.

Heckman, J. (1997), "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations", *Journal of Human Resources*, 32(3), 441-462.

_____. (2001), "Microdata, Heterogeneity and Econometric Policy Evaluation", Nobel Memorial Lecture in Economic Sciences, *Journal of Political Economy*, 109(4), 673-748.

Heckman, J., L. Lochner and P. Todd. (2003) "Fifty Years of Mincer Earnings Functions," Unpublished manuscript. University of Chicago.

Heckman, J. and S. Navarro-Lozano. (2003), "Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models" forthcoming in *Review of Economics and Statistics*. October.

Heckman, J. and E. Vytlacil (1999), "Local Instrumental Variable and Latent Variable Models for Identifying and Bounding Treatment Effects", *Proceedings of the National Academy of Sciences*, 96:4730-4734.36.

_____. (2000), "Local Instrumental Variables", in C. Hsiao, K. Morimune, and J. Powells, (eds.), *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya,* Cambridge: Cambridge University Press, 1-46.

_____. (2001), "Structural Equations, Treatment Effects and Econometric Policy Evaluation", Fisher Schultz Lecture World Congress of the Econometric Society.

National Bureau of Statistics (NBS). (2001), *China Labor Statistical Yearbook*. Beijing: China Statistics Press.

Navarro-Lozano, S. (2002), "Matching, Selection and the Propensity Score: Evidence from Training in Mexico," Working paper, presented at the Seventh Annual Meeting of LACEA, Madrid, Spain, October.

Table 1. Average Resident Income of Urban China in 2000

(in RMB yuan)

Provinces	Average Resident Income
Beijing	10350
Guangdong	9762
Zhejiang	9279
Sichuan	5894
Liaoning	5358
Shaanxi	5124
Average of the six provinces	7627
China average	6280

Source: NBS (2001), China Statistical Yearbook on Price & Urban Household Income and Expenditure Survey 2000, China Statistics Press, Beijing.

Table 2. Summary Statistics

Variable	All (n=587)		Treated (n=273)		Untreated (n=314)	
	Mean	Std. Err	Mean	Std. Err	Mean	Std. Err
Log Wage	8.86	0.86	9.12	0.77	8.64	0.88
Age	26.25	4.72	26.48	4.14	26.06	5.16
Years of work experience	6.41	4.92	5.83	4.47	6.91	5.23
4-Year college attendance	0.47	0.50	1	0	0	0
Male	0.56	0.50	0.54	0.50	0.59	0.49
Lived in Guangdong Province (GD)	0.18	0.39	0.19	0.39	0.18	0.38
Lived in Liaoning Province (LN)	0.28	0.45	0.30	0.46	0.27	0.44
Lived in Shaanxi Province (SX)	0.10	0.30	0.08	0.27	0.12	0.33
Lived in Sichuan Province (SC)	0.16	0.37	0.15	0.36	0.17	0.38
Lived in Beijing (BJ)	0.15	0.36	0.15	0.36	0.14	0.35
Lived in Zhejiang Province (ZJ)	0.12	0.33	0.12	0.33	0.12	0.33
Worked in state owned enterprises (SOEs)	0.62	0.49	0.72	0.45	0.54	0.50
Worked in collective-owned firms	0.08	0.27	0.04	0.20	0.11	0.32
Worked in joint-venture or foreign owned firms	0.18	0.39	0.19	0.40	0.17	0.38
Worked in private owned firms	0.12	0.32	0.05	0.21	0.18	0.38
Worked in IND_CON sector*	0.26	0.44	0.21	0.40	0.32	0.47
Worked in TRA_COM sector*	0.03	0.17	0.03	0.17	0.03	0.18
Worked in HOU_RES sector*	0.08	0.27	0.07	0.26	0.09	0.29
Worked in SPO_SOC sector*	0.22	0.41	0.16	0.36	0.27	0.45
Worked in CUL_SCI sector*	0.10	0.29	0.14	0.34	0.06	0.24
Worked in FIN_INS sector*	0.11	0.32	0.09	0.28	0.13	0.34
Worked in GOVERN sector*	0.03	0.16	0.04	0.20	0.02	0.13
Worked in OTHER sector*	0.17	0.38	0.27	0.45	0.08	0.28
Years of father's education	11.36	3.38	12.26	3.26	10.57	3.28
Years of mother's education	9.90	2.99	10.41	3.31	9.46	2.60
Parental income (in 1000 yuan)	21.39	16.59	24.36	15.89	18.81	16.78
Born before 1964	0.03	0.17	0.02	0.15	0.04	0.20
Born in 1964	0.02	0.14	0.01	0.10	0.03	0.17
Born in 1965	0.03	0.16	0.04	0.20	0.02	0.13
Born in 1966	0.02	0.14	0.03	0.16	0.02	0.13
Born in 1967	0.01	0.09	0.01	0.09	0.01	0.10
Born in 1968	0.03	0.16	0.03	0.17	0.02	0.15
Born in 1969	0.03	0.17	0.03	0.16	0.04	0.18
Born in 1970	0.05	0.22	0.03	0.16	0.07	0.26
Born in 1971	0.06	0.23	0.07	0.26	0.04	0.20
Born in 1972	0.05	0.22	0.05	0.22	0.05	0.21
Born in 1973	0.08	0.27	0.08	0.28	0.07	0.26
Born in 1974	0.09	0.28	0.10	0.30	0.07	0.26
Born in 1975	0.09	0.28	0.11	0.31	0.07	0.26
Born in 1976	0.11	0.31	0.14	0.35	0.08	0.27
Born in 1977	0.10	0.30	0.11	0.31	0.09	0.29
Born in 1978	0.10	0.30	0.11	0.31	0.09	0.29
Born in 1979	0.05	0.20	0.03	0.16	0.07	0.25
Born in 1980	0.04	0.20	0.01	0.10	0.07	0.25

*: IND_CON stands for the sectors of industry, geological exploration & census, and construction; TRA_COM for sectors of traffic, transportation, post and telecommunication, commerce, catering trade, and material supply; HOU_RES for sectors of housing & public utility management, and resident service; SPO_SOC for sectors of sanitation, sports, and social welfare; CUL_SCI for sectors of culture, arts, & education, science, research, and technology services; FIN_INS for sectors of finance and insurance; GOVERN for sectors of state and institutions, party and government mass organization; OTHER for all the other sectors.

Table 3. Estimated Mincer Model

Variable	OLS		IV	
	Coefficient	**Standard Error**	**Coefficient**	**Standard Error**
Intercept	8.3189	0.1493	8.3040	0.1552
4-Year's college attendance	0.2929	0.0630	0.5609	0.1695
Years of work experience	0.0380	0.0194	0.0196	0.0202
Experience squared	-0.0016	0.0010	-0.0007	0.0010
Parental income in 1000 yuan	0.0117	0.0020	0.0098	0.0023
Male	0.1537	0.0602	0.1439	0.0607
Lived in Guangdong Province	0.7543	0.1255	0.7908	0.1267
Lived in Liaoning Province	0.2693	0.1085	0.3142	0.1092
Lived in Sichuan Province	0.2278	0.1181	0.2759	0.1192
Lived in Beijing	0.7246	0.1241	0.7775	0.1256
Lived in Zhejiang Province	0.6241	0.1297	0.6739	0.1314
Worked in state owned enterprises	-0.3679	0.0855	-0.3873	0.0868
Worked in collective-owned firms	-0.4786	0.1288	-0.5890	0.1298
Worked in private owned firms	-0.4649	0.1179	-0.5304	0.1179
Worked in IND_CON sector*	-0.2793	0.0788	-0.3048	0.0792
Worked in TRA_COM sector*	-0.4512	0.1762	-0.4645	0.1779
Worked in SPO_SOC sector*	-0.2880	0.0900	-0.3106	0.0905
Worked in FIN_INS sector*	-0.3220	0.1050	-0.3327	0.1061

*: IND_CON stands for the sectors of industry, geological exploration & census, and construction; TRA_COM for sectors of traffic, transportation, post and telecommunication, commerce, catering trade, and material supply; SPO_SOC for sectors of sanitation, sports, and social welfare; FIN_INS for sectors of finance and insurance.

#: Using Propensity score as the instrument for four-year college attendance (Instruments are parental education and year of birth).

Table 4. Estimated Logit Model For Schooling

Variable	Coefficient	Standard Error	Mean Marginal Effect
Intercept	-4.7370	0.7305	-
Years of father's education	0.1017	0.0297	0.0211
Years of mother's education	0.0605	0.0342	0.0126
Parental income in 1000 yuan	0.0190	0.0069	0.0040
Born before 1964	2.0008	0.7969	0.4159
Born in 1964	1.7285	0.9189	0.3593
Born in 1965	3.3423	0.8257	0.6947
Born in 1966	3.1813	0.8552	0.6613
Born in 1967	1.8455	1.1126	0.3836
Born in 1968	2.9030	0.8161	0.6034
Born in 1969	2.2569	0.7941	0.4691
Born in 1970	1.5076	0.7534	0.3134
Born in 1971	3.0771	0.7138	0.6396
Born in 1972	2.6424	0.7183	0.5492
Born in 1973	2.5395	0.6809	0.5279
Born in 1974	2.7740	0.6753	0.5766
Born in 1975	2.7931	0.6763	0.5806
Born in 1976	2.8634	0.6669	0.5952
Born in 1977	2.5890	0.6672	0.5381
Born in 1978	2.5572	0.6656	0.5315
Born in 1979	1.3631	0.7636	0.2833

Table 5. Estimated Coefficients from Local Linear Regression

Guassian Kernel, bandwidth = 0.4

Variable	High School		College	
	γ_0	**Std. Err.**	γ_1	**Std. Err.**
Years of work experience	0.0360	0.0225	0.0141	0.0278
Experience squared	-0.0013	0.0011	-0.0009	0.0013
Parental income in 1000 yuan	0.0188	0.0038	0.0077	0.0038
Male	0.1365	0.0723	0.1913	0.0777
Lived in Guangdong Province	0.5712	0.1961	0.8853	0.1590
Lived in Liaoning Province	0.1901	0.1263	0.3929	0.1049
Lived in Sichuan Province	0.2612	0.1364	0.2296	0.1081
Lived in Beijing	0.7122	0.1695	0.7971	0.1301
Lived in Zhejiang Province	0.6930	0.1551	0.5461	0.1744
Worked in state owned enterprises	-0.3368	0.1188	-0.4471	0.1093
Worked in collective-owned firms	-0.6060	0.2065	-0.5868	0.1771
Worked in private owned firms	-0.4205	0.1511	-0.6256	0.1677
Worked in IND_CON sector*	-0.2297	0.0821	-0.3978	0.0990
Worked in TRA_COM sector*	-0.3527	0.1318	-0.5040	0.1557
Worked in SPO_SOC sector*	-0.3702	0.1282	-0.3040	0.1202
Worked in FIN_INS sector*	-0.3345	0.1560	-0.3543	0.1331

*: IND_CON stands for the sectors of industry, geological exploration & census, and construction; TRA_COM for sectors of traffic, transportation, post and telecommunication, commerce, catering trade, and material supply; SPO_SOC for sectors of sanitation, sports, and social welfare; FIN_INS for sectors of finance and insurance.

Table 6. Comparison of Different Parameters

Parameter	Estimation
OLS	0.2929
IV*	0.5609
ATE	0.4336
TT	0.5149
TUT	0.3630
Bias**	-0.1407
Selection Bias***	-0.2220
Sorting Gain****	0.0813

* Using propensity score as instrument

** $Bias = OLS - ATE$

*** $Selection\ Bias = OLS - TT$

**** $Sorting\ Gain = TT - ATE$

Table 7a. Estimates of Returns to schooling

Including both firms' ownership and sectors' dummies in wage equation

Parameter	Annualized	
	With parental income	Without parental income
OLS	0.0732	0.0856
IV*	0.1402	0.2192
ATE	0.1084	0.2321
TT	0.1287	0.1909
TUT	0.0908	0.2679

*Using propensity score as the instrument

Table 7b. Estimates of Returns to schooling

Including firms' ownership dummies only in wage equation

Parameter	Annualized	
	With parental income	Without parental income
OLS	0.0873	0.1010
IV*	0.1777	0.2549
ATE	0.1439	0.2659
TT	0.1588	0.2199
TUT	0.1309	0.3059

*Using propensity score as the instrument

Table 7c. Estimates of Returns to schooling

Including sectors' dummies only in wage equation

Parameter	Annualized	
	With parental income	Without parental income
OLS	0.0802	0.0920
IV*	0.1241	0.2049
ATE	0.0960	0.2204
TT	0.1059	0.1639
TUT	0.0874	0.2694

*Using propensity score as the instrument

Table 7d. Estimates of Returns to schooling

Excluding both firms' ownership and sectors' dummies in wage equation

Parameter	Annualized	
	With parental income	Without parental income
OLS	0.0968	0.1109
IV*	0.1703	0.2496
ATE	0.1420	0.2630
TT	0.1438	0.1992
TUT	0.1404	0.3185

*Using propensity score as the instrument

Figure 1. Density of P(S=1)
Urban areas of six provinces of China
From CUHIES 2000

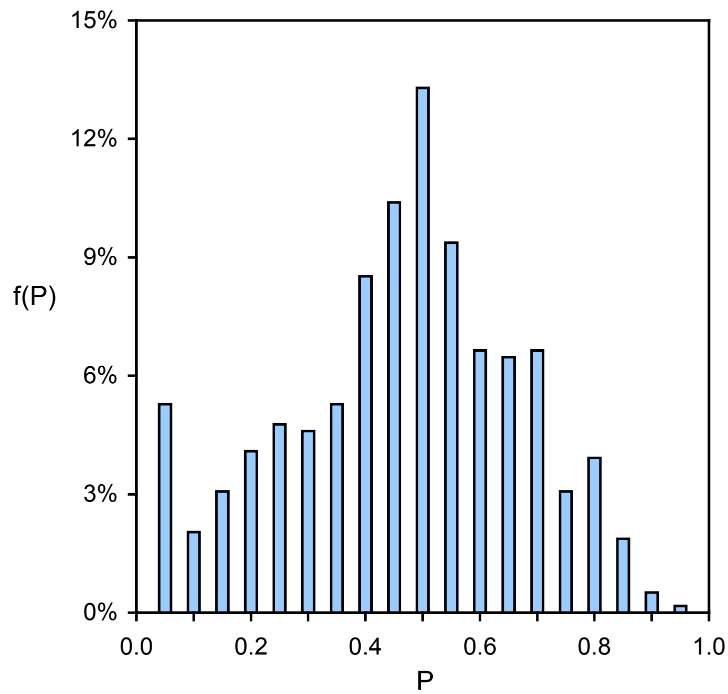


Figure 2. Marginal Treatment Effect
Including parental income as proxy for ability
in wage equation, all ownership and sectoral dummies
also included, Bandwidth = 0.4

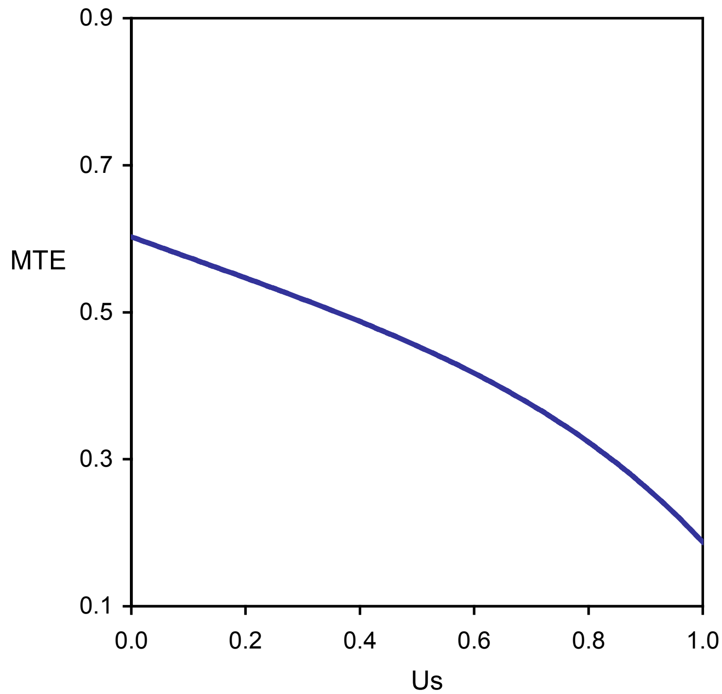


Figure 3. Weights of Treatment Parameters For Main Specifications

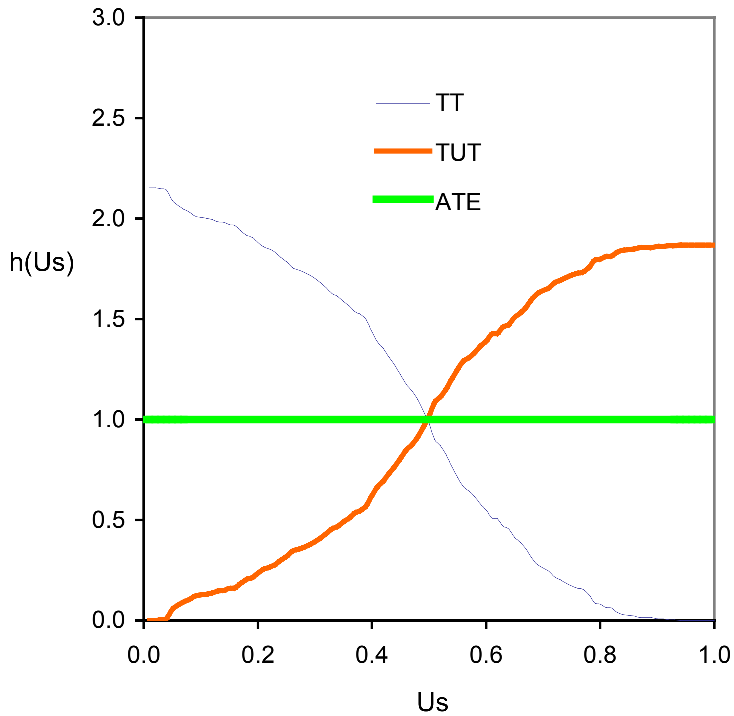


Figure 4. Marginal Treatment Effect

Excluding parental income in wage equation
But all ownership and sectoral dummies included
Bandwidth = 0.3

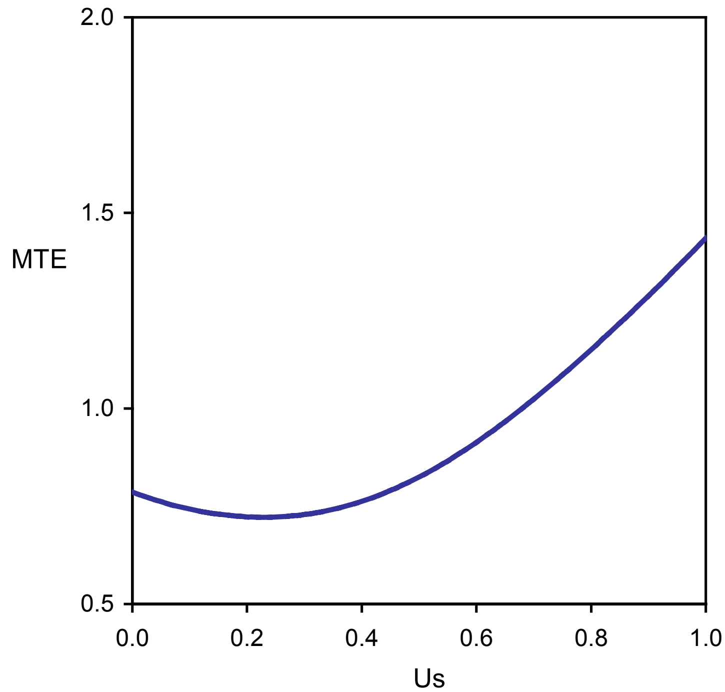
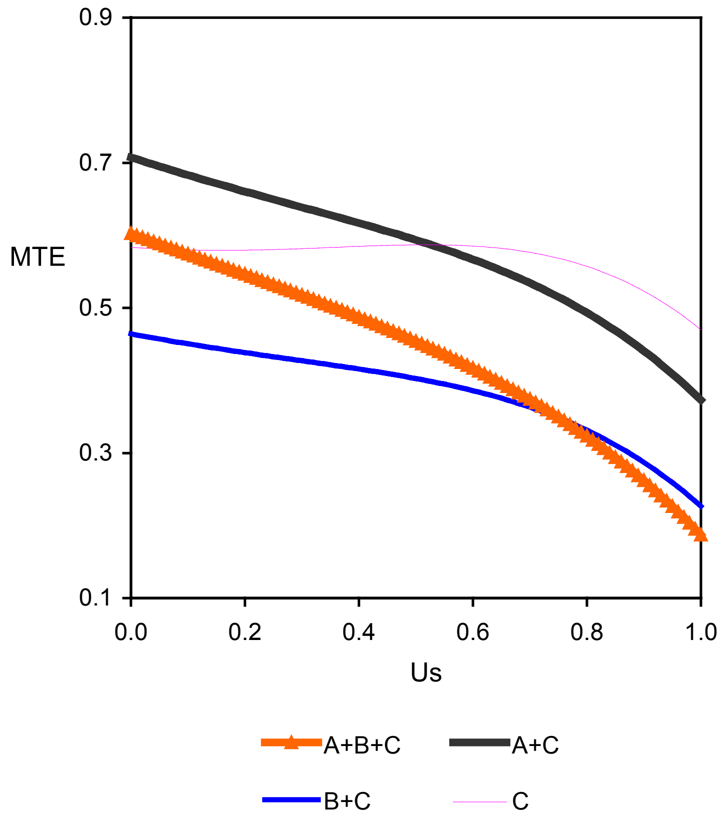


Figure 5. Marginal Treatment Effect
All specifications include parental income
in earnings equation, Bandwidth = 0.4



A: with firms' ownership dummies but not sectoral dummies

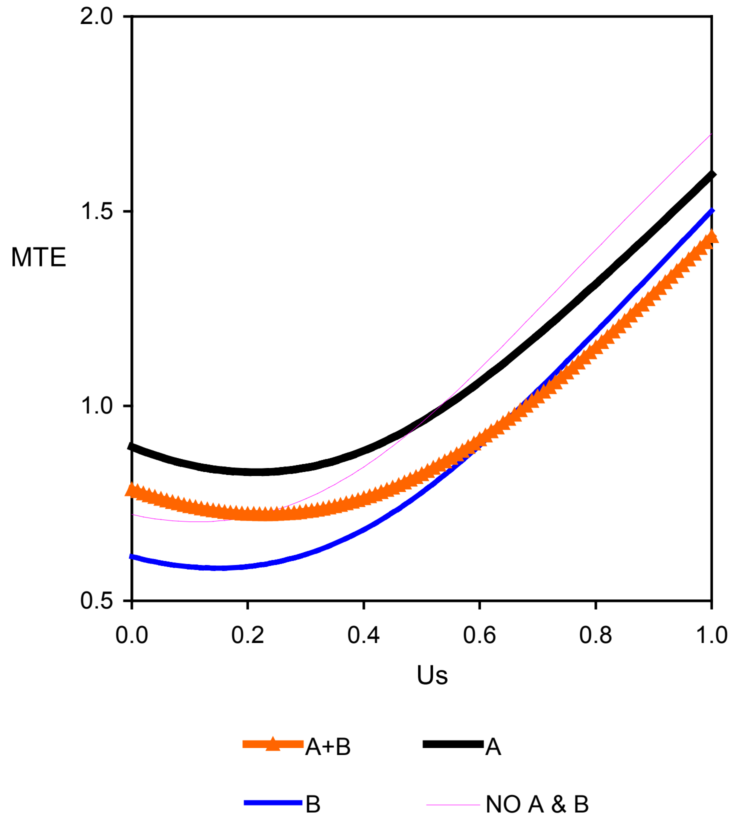
B: with sectoral dummies but not ownership dummies

C: with parental income as proxy for ability,

no sectoral and ownership dummies

Figure 6. Marginal Treatment Effect

All specifications exclude parental income in earnings equation, Bandwidth = 0.3



A: only with firms' ownership dummies

B: only with sectoral dummies