Access to Higher Education and Inequality: The Chinese Experiment

May 2007

Xiaojun Wang
Department of Economics
University of Hawaii at Manoa
2424 Maile Way 527
Honolulu, HI 96822
Email: xiaojun@hawaii.edu

Belton M. Fleisher
Department of Economics
The Ohio State University
1945 N. High St.
Columbus, OH 43210
IZA
Email: fleisher.1@osu.edu

Haizheng Li
Department of Economics
Georgia Tech
Atlanta, Georgia

Shi Li
Chinese Academy of Social Sciences
Beijing, China

We are grateful to Pedro Carneiro, Joe Kaboski, James Heckman, and Edward Vytlacil for their invaluable help and advice and to Sergio Urzúa for providing help and advice with software codes. Quheng Deng contributed invaluable research assistance.

Access to Higher Education and Inequality: The Chinese Experiment

Abstract — We apply a semi-parametric latent variable model to estimate selection and sorting effects on the evolution of private returns to schooling for college graduates during China's reform between 1988 and 2002. We find that there were substantial sorting gains under the traditional system, but they have decreased drastically and become negligible in the most recent data. We take this as evidence of growing influence of private financial constraints on decisions to attend college as tuition costs have risen and the relative importance of government subsidies has declined. The main policy implication of our results is that labor and education reform without concomitant capital market reform and government support for the financially disadvantaged exacerbates increases in inequality inherent in elimination of the traditional "wage-grid."

Keywords: Return to schooling, selection bias, sorting gains, heterogeneity, financial constraints, comparative advantage, China

JEL Codes: J31, J24, O15

I. Introduction and Background

Two salient features of the labor force in centrally planned economies were the wage-grid and the *nomenklatura*. The wage-grid system compressed wage differentials across education groups, while the *nomenklatura* system selected who attended college to acquire knowledge and training to function in the planning bureaucracy. Consequently, higher educational attainment was not an outcome of free choice and the economic return to higher education in terms of wages tended to be very low. Since 1978 (China) and approximately 1990 (the former Soviet Union and its satellites), most of the world's planned economies have abolished central planning and have entered a period of transition to market systems. During transition, wage-grids have been relaxed or removed, and wage differentials increasingly reflect the market outcomes; educational attainment, especially at higher levels, has become subject to conscientious choices made by each individual; conventionally estimated returns to education have risen to levels comparable to those observed in developed countries. However, transition toward free markets has occurred at different speeds across the formerly planned economies, and wage differential trajectories have varied widely.[1]

Among the major transitional economies, China has taken the most gradual course toward market reform. From the inception of economic reform into the early 1990s, wage differences by level of skill, occupation, and/or schooling remained very narrow in China. The Mincerian return to higher education was even lower than that in the early years of the Mao era (Fleisher and Wang, 2005). Since the early 1990s, there is evidence that returns to schooling in China have begun to increase (Zhang and Zhao, 2002; Li, 2003; Yang, 2005). Although a rising return to schooling has probably contributed to growing income inequality,[2] it is our view that *access* to education is a more important

[1] There is a growing literature on returns to education and wage differentials experienced in these transitional economies. See Brainerd (1998) on Russia; Munich, Svejnar and Terrell (2005) on the Czech Republic; Orazem and Vodopivec (1995) on Slovenia; and Jones and Ilayperuma (2005) on Bulgaria. Fleisher, Sabirianova and Wang (2005) provide a comparative study of eleven former centrally planned economies including Russia and China.

[2] Yang (2005) shows that the dispersion of returns to schooling across Chinese cities increased sharply between 1988 and 1995. Wang et al (2007) provide most recent evidence of rising income inequality from 1987 to 2002 in China.

factor. According to Yang (1999), China in the late 1990s surpassed almost all countries in the world for which data are available in rising income inequality, and by the year 2000 China found itself with one of the highest degrees of income inequality in the world (Yang, 2002).

We are concerned with the question of whether rising inequality in China is associated with disparate access to educational opportunities. The end of the Mao era saw the influence of political considerations on access to higher education sharply diminish, and college admission criteria reverted to historical practice which placed a very heavy weight on merit as determined by critical tests in senior high schools. More recently, however, a growing proportion of college students must fund their own educational expenses (Hannum, 2004; Heckman, 2004), and many college-worthy students are shut out due to financial considerations.[3] Between 1992 and 2003, the proportion of government expenditures in total education expenditures in China decreased from 84% to 62%, and the proportion of tuition and fees increased from about 5% to approximately 18% (*China Statistical Yearbook* 2005). The proportion of the population privileged to attend college has been and remains very small by almost any standard, despite a sharp acceleration of schooling expenditures and expansion of enrolment in the past decade (Fleisher and Wang, 2005; Heckman, 2005). The proportion of college graduates in the total population was 0.6% in 1982, 1.4% in 1990, 2.0% in 1995, 4.1% in 2001, and 5.2% in 2003, according to various issues of *China Statistical Yearbook*.

Access to college and concomitant economic gain depends not only on current financial resources, but also on the ability to achieve high test scores and on cognitive and other attributes produced in earlier family and educational contexts. Thus, higher educational attainment depends recursively on earlier access to publicly and privately supported education at lower levels as well as on the capacity to borrow funds from family and other sources to pay direct and indirect college costs (Carneiro and Heckman, 2002; Hannum, 2004). If access to all levels of schooling is available only to the

---

[3] Until the early 1990s, college education was almost free in China. The government paid for tuition and lodging, while students only needed to pay for meals and books. Tuition at major Chinese Universities now approaches US$1,000 per year or more (People's Daily, 2000).

financially, politically, and geographically advantaged, the bulk of China's population will be excluded from full participation in the growth of human capital and the income it produces.

In this paper we focus on the *changes* in returns to college education in China from 1988 to 2002, paying particular attention to sorting and selection under heterogeneous returns. We address the following questions.

1. How has the relative importance of variables that determine the probability of college attendance changed?
2. Is there evidence that the degree of purposive selecting the college alternative over stopping with a high school diploma has changed?
   - Has the gain from choosing college narrowed or widened?
   - If it has widened, is this because more qualified students are now able to attend college due to reduced favoritism?
   - If it has narrowed, is this consistent with an efficient process with an increased proportion of qualified college graduates graduating from college?
   - Is there evidence of increased influence of borrowing constraints, which would prevent high-school graduates from choosing the college alternative even though they would reap returns as high as or even higher than those who do graduate from college?

Our major contribution is to estimate both the levels of and *changes* in the returns to a four-year college education over a critical time period of China's transition. During the period covered by our data China moved from economic planning and a regime of education choices that were strictly prescribed and paid for by the planning regime to a market economy and a regime in which individuals and their families are increasingly free to make their own decisions. But "free to choose" has increasingly meant "required to pay," as well, and students and their families must now finance an increasing proportion of the cost of higher education. The literature has largely ignored the impact of lagging capital market reform on individual investment in human capital. In this paper, we shed some light on the effects of this lack of coordination in reform.

We exploit three cross-sectional data sets of 1988, 1995, and 2002, respectively.[4] Since 1989, China's higher education began to transform from tuition-free (with some living allowances to students) to almost fully privately funded. By 1997, tuition had become mandatory in all colleges in China. Our three sample years represent these distinct stages nicely: 1988 is the antecedent stage when government still subsidized almost all the tuition cost of higher education; 1995 was in the midst of the transition to a more private-funded system; by 2002, the transition was well advanced. Therefore, this multi-year data set allows us to examine how this policy change has affected individual choices and outcomes in higher education. Particularly, this policy change has raised concerns that some college-worthy youth may not be able to attend due to borrowing constraints.[5] This inefficiency in the education system not only implies loss of future productivity, it is also likely to exacerbate income disparity. Therefore, by comparing estimates from before, during, and after this major transformation, we are able to assess the merits and shortcomings of this profound policy change.

We use methods developed in Heckman and Vytlacil (1999, 2000) that combines the treatment effect literature (Bjorklund and Moffitt, 1987) with the latent variable literature. Griliches (1977) considers a model with homogeneous returns in which unobserved ability and measurement error pose the major threats to estimation. Therefore, instrumental variable (IV) is suggested to correct bias in the estimators. However, this solution breaks down when one follows the other strand of research pioneered by Roy (1951), Willis and Rosen (1979), and Willis (1986). These scholars assume that schooling decisions are conscientious choices by rational forward-looking individuals who are capable of at least partially anticipating the return to education and that they act upon it. Therefore, the appropriate method is to estimate a latent variable model with correlated random coefficients.

Heckman and Vytlacil (1999, 2000) and Caneiro, Heckman, and Vytlacil (2000) explain why conventional approaches fail to estimate meaningful policy parameters when there are heterogeneous returns in the population and people act upon them. Suppose the return to schooling parameter, β, is randomly distributed across the population as shown

[4] These are not panel data sets.
[5] Consumption credit, including student loans, was still rare during our sample period.

in Figure 1. Ignoring the heterogeneity and uncertainty in the costs of attaining education, let β_1 be the current cut-off return. That is, only those whose return to education is greater than β_1 will find it worthwhile to attend school. There are several interesting policy parameters in this framework, but it is unclear which one the conventional instrumental variable method estimates. For example, the mean return for those who attend school is $\int_{\beta_1}^{\infty} \beta dF(\beta)$ where $F(\beta)$ is the cumulative distribution function of the returns to education, the mean return for those who do not attend school is $\int_{-\infty}^{\beta_1} \beta dF(\beta)$, and the population mean return is $\int_{-\infty}^{\infty} \beta dF(\beta)$. Suppose a tuition hike pushes the cut-off return up to β_2, then the conventional instrumental variable method estimates $\int_{\beta_1}^{\beta_2} \beta dF(\beta)$ — the average return of those whose schooling decisions are reversed due to the tuition hike, which in general doesn't agree with any of the above policy parameters.[6] That is, the conventional instrumental variable method doesn't recover appropriate policy parameters because the subset of returns of those who reverse decisions due to the instruments is not representative of the schooled, the unschooled, or the population as reflected in the entire hypothetical distribution of returns depicted in Figure 1. In this paper we estimate parameters that answer well-posed policy questions.

The rest of the paper is organized as follows. Section II presents the theoretical framework and derives parameters that answer well-posed policy questions. Section III briefly discusses the data. Empirical results are reported and analyzed in section IV. Section V draws conclusion.

II. Methodology

Our method takes into account both heterogeneous returns to schooling and self-selection based on anticipated returns. We first estimate the marginal treatment effect

---

[6] The same also applies to other popular instrument variables used in the literature such as compulsory schooling and distance to nearest schools, etc.

(MTE) in the samples, which is the building block of other parameters of interest.[7] The marginal treatment effect and parameters derived from it are estimated using the *local instrumental variable* method developed in Heckman, Ichimura, Todd, and Smith (1998).[8]

We set up the following model of earnings determination by schooling choice:

$$\begin{aligned} \ln Y_1 &= \mu_1(X) + U_1 \\ \ln Y_0 &= \mu_0(X) + U_0 \end{aligned} \qquad (1)$$

where a subscript indicates whether the individual is in the schooled state (S=1) or the unschooled state (S=0).[9] Y is a measure of income, and X is observed heterogeneity that might explain earnings differences. U_1 and U_0 are unobserved heterogeneities in earnings determination, and $E(U_1)=E(U_2)=0$. In general, the functional forms can have a nonlinear component, and $U_1 \neq U_0$.

Each individual can choose only one of the above two states. The schooling choice comes from the following latent variable model:

$$\begin{aligned} &S^* = \mu_s(Z) - U_s \\ &S = 1 \;\; if \;\; S^* \geq 0 \\ &S = 0 \;\; otherwise \end{aligned} \qquad (2)$$

where S^* is a latent variable whose value is determined by an observed component $\mu_s(Z)$ and a unobserved component U_s. A rational individual will attend college (i.e. S=1) only if this latent variable is nonnegative.

---

[7] Marginal treatment effect is the marginal gain to schooling of a person just indifferent between taking schooling or not. See Bjorklund and Moffitt (1987) and Carneiro, Heckman, and Vytlacil (2000).
[8] These derivatives include average treatment effect (ATE), treatment on the treated (TT), treatment on the untreated (TUT), bias, selection bias, and sorting gain. Each of them is defined explicitly below.
[9] Throughout this paper the schooled state is attending college, while the unschooled state is not attending college after graduating from high school. Sometimes the college state is also referred to as the treated state, while high-school graduates are sometimes referred to as the untreated state. We only consider individuals who at least have graduated from high school.

In our empirical work, Z is a vector of variables that helps predict the probability of attending college. It includes parental education, parental income, number of siblings, gender, ethnicity, and birth year dummies. X is a vector of variables that helps explain earnings. In the benchmark model, X includes work experience, work experience squared, gender, ethnicity, and three firm-level characteristics: ownership, industry, and location. Z and X can share some common variables, but Z must also possess unique variables for the model to be identified. That is, variables included in Z but not in X serve as instruments to identify the returns to education, and these instruments are applied *locally* so that they identify each region in the distribution of the marginal treatment effects.[10] It is obvious that equations (1) and (2) are correlated not only because X and Z usually share components, but also because the schooling decision is at least partially made based on anticipation of the returns implied in the potential earnings equations.

College entrance in China has been highly competitive since its resumption in 1979. Only a small fraction of high school graduates can pass the rigorous National College Entrance Exams and continue the pursuit of higher education. Moreover, students have been required to pay at least part of their tuition since the early 1990s, which has made college attendance more difficult for financially disadvantaged families. In estimating the schooling choice model, we use both parental income and parental education to control for ability formation and possible financial constraints. Research on human resources is abundant with evidence that children from well-educated parents are more likely to go to college. Higher parental income not only mitigates short-run financial constraints, it also predicts long-run ability-enhancing benefits due to better earlier education, better nutrition, and better environments that foster cognitive and non-cognitive skills in children. The change in policy on public versus private financing of higher education offers a rare opportunity to analyze how the roles played by parental income and education have changed.

Equations (1) define a heterogeneous return to education,

---

[10] The conventional *global* instrument method (see Figure 1) only identifies the mean return of the subset of people whose decisions are reversed by the instrument. However this subset does not in general represent the either treated, the untreated, or the population. By applying an instrument locally, we circumvent the "representativeness" issue by identifying a limiting version of the return, i.e. the marginal treatment effect.

$$\beta = \ln Y_1 - \ln Y_0 = \left(\mu_1(X) - \mu_0(X)\right) + \left(U_1 - U_0\right) \tag{3}$$

Therefore β is a random variable that is correlated with U_0 and U_1. Pooling the schooled and unschooled together,

$$\ln Y = S \ln Y_1 + (1-S)\ln Y_0 = \ln Y_0 + \beta S = \mu_0(X) + \beta S + U_0 \tag{4}$$

Equations (3) and (4) reveal the problems in conventional OLS estimation. More specifically, Heckman and Li (2004) shows[11]

$$\begin{aligned} p\lim\left(\hat{\beta}_{OLS}\right) &= E\left(\ln Y_1 \mid S=1\right) - E\left(\ln Y_0 \mid S=0\right) \\ &= E\left(\mu_1 - \mu_0\right) + \left[E\left(U_1 \mid S=1\right) - E\left(U_0 \mid S=0\right)\right] \end{aligned} \tag{5}$$

The first term is the average treatment effect (ATE), i.e. the rate of return to education for a randomly selected individual. The second term in the square bracket is the OLS *bias* and it can be either positive or negative. Therefore, OLS in general doesn't estimate the average treatment effect consistently. From the perspective of individuals who choose college, the OLS bias can be decomposed as follows:

$$\begin{aligned} &E\left(U_1 \mid S=1\right) - E\left(U_0 \mid S=0\right) \\ &= \left[E\left(U_0 \mid S=1\right) - E\left(U_0 \mid S=0\right)\right] + E\left(U_1 - U_0 \mid S=1\right). \end{aligned} \tag{6}$$

From the perspective of the unschooled group, the decomposition of the OLS bias is:

$$\begin{aligned} &E\left(U_0 \mid S=0\right) - E\left(U_1 \mid S=1\right) \\ &= \left[E\left(U_1 \mid S=0\right) - E\left(U_1 \mid S=1\right)\right] + E\left(U_0 - U_1 \mid S=0\right) \end{aligned} \tag{7}$$

The term in the square bracket in (6) is *selection bias* for college students. It is the mean difference in unobservables between the counterfactual of what a college graduate would

---

[11] We suppress the conditioning of X here and below in order to simplify exposition.

earn if he didn't attend college and what an average high school graduate earns. The next term is *sorting gain*, which is the mean gain in the unobservables for college graduates, i.e. the counterfactual difference between what an average college graduate earns and what he would earn if the college degree were not obtained. In Equation (7), the bracketed term is the selection bias for the unschooled group, which is the mean difference in unobservables between the counterfactual of what a high school graduate would earn had he completed college and what an average college graduate earns. The second term is the sorting gain for this group, which is the mean difference in unobservables for high school graduates, i.e. the difference between what an average high school graduate earns and the counterfactual of what would be earned had he completed college.[12]

Willis and Rosen (1979) show that selection biases — the first terms in equations (6) and (7) — can be either positive or negative. When they are both negative, it is consistent with sorting by comparative advantage. On the other hand, positive selection bias in equation (6) and negative selection bias in equation (7) would be consistent with a single-factor (hierarchical) interpretation of ability, i.e. the schooled group on average has higher ability than the unschooled group.[13]

Combine the above two types of sorting gains with the average treatment effect, we obtain two parameters that are of great policy interest:

$$\begin{aligned} E(\beta \mid S=1) &= E(\ln Y_1 - \ln Y_0 \mid S=1) = E(\beta) + E(U_1 - U_0 \mid S=1) \\ E(\beta \mid S=0) &= E(\ln Y_1 - \ln Y_0 \mid S=0) = E(\beta) - E(U_0 - U_1 \mid S=0) \end{aligned} \tag{8}$$

The first equation defines the treatment on the treated effect (TT), and it can be decomposed into the sum of the average treatment effect and the sorting gain for the schooled group. The second equation defines the treatment on the untreated effect (TUT),

[12] Selection bias compares two groups of persons, the schooled and unschooled, while sorting gain compares two distinct earning results of the same group. Therefore, the above decompositions by group allow us to extract much more information from the data than most conventional methods.

[13] For example, if the labor market is dominated by sorting by comparative advantages, then the best lawyers (i.e. schooled or college graduates) are also the worst plumbers (i.e. unschooled or high school graduates), and vice versa. Under the hierarchical ability assumption, however, typical college graduates would be more productive lawyers *and* plumbers than typical high school graduates.

which is the average treatment effect minus the sorting gain for the unschooled group. The treatment on the treated effect captures the mean gain the schooled group experience, compared with what they would earn if they hadn't gone to college. The treatment on the untreated effect captures the mean gain the unschooled group would experience if they had gone to college, compared with what they earn now. If the sorting gain for the schooled group is positive, it is evidence of purposive sorting based on heterogeneous returns to education. It is particularly interesting to note that a negative sorting gain for the unschooled group signals possible involuntary sorting, meaning that the unschooled group would prefer to go to college, but are restrained from selecting their preferred alternative by unobserved barriers to college attendance.

Selection bias can be obtained from the following alternative decomposition of the OLS estimator:

$$\begin{aligned} p\lim\left(\hat{\beta}_{OLS}\right) &= E\left(\ln Y_1 \mid S=1\right)-E\left(\ln Y_0 \mid S=0\right) \\ &= E\left(\beta \mid S=1\right)+\left[E\left(U_0 \mid S=1\right)-E\left(U_0 \mid S=0\right)\right] \\ &= E\left(\beta \mid S=0\right)-\left[E\left(U_1 \mid S=0\right)-E\left(U_1 \mid S=1\right)\right] \end{aligned} \quad (9)$$

Tautologically, the selection bias for the schooled group is the difference between the OLS estimate and treatment on the treated effect, while the selection bias for the unschooled group is the difference between the treatment on the untreated effect and the OLS estimate.

Following Carneiro, Heckman, and Vytlacil (2000), we adopt a two-step procedure to estimate the above parameters. In the first step, a probit model is used to estimate the $\mu_s(Z)$ function of equation (2). The predicted value is called the propensity score, $\hat{P}_i$, where the subscript *i* denotes each individual. The second step adopts a semi-parametric procedure in which *local polynomial regressions* are used to retrieve the marginal treatment effect. That is, we do not impose any functional restrictions on the relation between marginal treatment effect and unobservables in the schooling choice equation. Fan (1992, 1993) develop the distribution theory for the local polynomial estimator of $E(\Phi|\Xi=\xi)$, where (Φ,Ξ) is a bivariate random data set. It is shown that

$E(\Phi|\Xi=\xi)$ and its first derivative can be consistently estimated by the following algorithm:

$$\min_{\gamma_1,\gamma_2}\sum_{i\le N}\left[\Phi_i-\gamma_1-\gamma_2\left(\Xi_i-\xi\right)\right]^2 G\left(\frac{\Xi_i-\xi}{a_N}\right) \qquad (10)$$

where N is the sample size. Then, γ_1 is a consistent estimator of $E(\Phi|\Xi=\xi)$, and γ_2 is a consistent estimator of $\partial E(\Phi|\Xi=\xi)/\partial\Xi$. $G(.)$ is a kernel function and a_N is a bandwidth. We use a Gaussian kernel and a bandwidth of 0.2 in the empirical estimation.[14] Intuitively, this algorithm is equivalent to applying weighted least squares at designated point, i.e. $\Xi=\xi$, using all observations but with decaying weights assigned to more distant data points.

More specifically, we estimate a partially linear, conditional expectation model of equation (3)

$$E\left(\beta\,|\,X,U_s=p\right)=\left(\mu_1\left(X\right)-\mu_0\left(X\right)\right)+E\left(U_1-U_0\,|\,X,U_s=p\right) \qquad (3')$$

By definition the left-hand-side is the marginal treatment effect at $U_s=\mu_s$. We assume linear functional forms for the first term on the right-hand-side of equation (3'), while we estimate the second term, i.e. $E(U_1-U_0|X,U_s=p)$ in a nonparametric manner.

Following the convention in the literature of semi-parametric estimation (Ichimura and Todd 2004), we first obtain consistent estimates of the linear coefficients with the *double residual regression*, and then retrieve the residuals for the nonparametric estimation. Specifically, we first estimate $E(lnY|P=p)$ and $E(X|P=p)$ with the local polynomial algorithm. Then we run the *double residual regression* of $lnY-E(lnY|P=p)$ on $X-E(X|P=p)$.[15] This is a simple OLS regression that yields consistent estimates of coefficients of the linear components of equation (1).[16] Let α be the vector of these

[14] This approximates the rule-of-thumb bandwidth selector proposed in Fan and Gilbels (1996).
[15] This procedure is analogous to de-meaning the earnings equations. This approximately purges out the nonlinear components due to the continuity of the nonlinear functions, which allows us to retrieve the nonlinear components later by using the residuals.
[16] A researcher can pick any reasonable set of propensity values to evaluate the conditional means and first derivatives. It is only bounded by the joint set of propensities for the schooled and unschooled groups. We use evenly spaced points from the joint set (with increment equal to 0.01).

estimates.[17] Define the nonlinear component residual as $U=lnY-\alpha X$. Use local polynomial regression again to estimate $E(U|P=p)$ and its first derivative. This first derivative by definition is the marginal treatment effect. The average treatment effect (ATE) is a simple integration (over the support of μ_s) of the MTE with equal weight assigned to each $U_s=\mu_s$. Furthermore, treatment on the treated (TT) and treatment on the untreated (TUT) are simple integration of MTE with the following weighting functions:[18]

$$\begin{aligned} h_{TT}(u_s) &= \frac{\left[\int_{u_s}^{1} f(p)\,dp\right]}{E(p)} \\ h_{TUT}(u_s) &= \frac{\left[\int_{0}^{u_s} f(p)\,dp\right]}{E(1-p)} \end{aligned} \tag{11}$$

where $f(p)$ is the conditional density of propensity scores. The conditioning on X is implicit in the above functions. All integrations are conducted numerically using simple trapezoidal rules.

The intuition behind this seemingly complex procedure is straightforward. Without loss of generality, assume the unobserved heterogeneity in the schooling choice function, U_s, follows a uniform distribution between 0 and 1.[19] To find out the marginal treatment effect for a given $U_s=\mu_s$, we apply the method of local polynomial regression, i.e. equation (10), to the linear and nonlinear components of equation (4) respectively. Since we are only interested in the marginal individuals whose unobserved heterogeneity of attending college is μ_s, a propensity score that is close to μ_s provides more information than that is farther away from μ_s. Because a Gaussian kernel is used in the empirical work

---

[17] Since we pool both the schooled and unschooled groups in this step, we essentially assume the linear components between the two equations in (1) only differ by a constant, which is the average treatment effect. This assumption dramatically simplifies the computation, and it can be easily modified by running separate double residual regression for each group and obtain different α for each group.
[18] For derivations of these weighting functions, see Heckman and Vytlacil (1999, 2000). The TT weight is basically the scaled probability of receiving a propensity score that is greater than μ_s, i.e. being treated. On the other hand, the TUT weight is the scaled probability of receiving a propensity score that is smaller than μ_s, i.e. not being treated.
[19] If this is violated, a simple transformation of the $P(Z)$ function can restore this assumption.

presented below, all observations are used in the estimation of the marginal treatment effect at each prescribed value of μ_s , but certainly those observations whose propensity scores are closer to μ_s dominate the estimates. Finally, since we are interested in the *change* in the logarithm of income, i.e. return to schooling, due to an infinitesimal change in μ_s, instead of the mean logarithm of income itself (γ_1 consistently estimates this value), the first derivative estimator γ_2 in equation (10) consistently estimates the marginal treatment effect.[20] Figure 2 provides an illustrative example on how marginal treatment effect is estimated. First pick a value of μ_s, the unobservable in schooling choice. The corresponding marginal treatment effect at this evaluation point is estimated with all observations. In figure 2 we pick two values of μ_s and eight propensity scores (i.e. eight observations; note it is the same eight observations that are used for each μ_s.). The 45-degree line on the μ_s-propensity plane identifies the points where the propensity score is equal to the unobserved μ_s, and according to the definition of marginal treatment effect this is where individuals are indifferent between the schooled and the unschooled states. We estimate the mean marginal return by exploiting the fact that some of these eight observations are associated with the schooled state while others with the unschooled state; and some are closer to the point of evaluation while others are farther away. This is estimated by using equation (10) with a Gaussian kernel. Suppose we use the vertical axis to represent the estimated values of marginal treatment effect, a projection onto the μ_s-MTE plane produces the entire curve of marginal treatment effects. Empirically we pick about one hundred evaluation points for μ_s rather than just two; greater accuracy can be achieved at the expense of longer computation time.

III. Data and Descriptive Statistics

The data used in this study are from the first, second, and third waves of the Chinese Household Income Project (CHIP) conducted in 1989 (CHIP-88), 1996 (CHIP-95), and 2003 (CHIP-2002). It was funded by a number of agencies and institutes, and

[20] In practice we also include a quadratic term in equation (10) to improve the accuracy of estimation.

was conducted by the Institute of Economics at the Chinese Academy of Social Sciences.[21] Each wave of the CHIP consists of an urban survey and a rural survey; we only use the urban survey data for this study. Each urban survey covers thousands of households and individuals in the urban area of about a dozen provinces. For example, CHIP-95 covers 6,928 households and 21,688 individuals in 11 provinces; and CHIP-02 covers 6,835 households and 20,632 individuals in 12 provinces.

These three sample years also represent distinct phases of economic reform in China. Specifically, 1988 represents the early stage of urban reform that started in 1982 and ended with the 1989 Tian-An-Men square demonstration. 1995 represents the middle stage of urban economic transitions after the reform re-started in 1992 and before the 1997-98 Asian financial crises. 2002 can be viewed as a relatively mature stage of economic transition. One measure of progress in economic transition is the share of employment in the non-public sector, because the ultimate goal of China's reform is to establish a market system. As reported in tables 2a, 2b and 2c, this proportion increased from 1% in 1988, to 9% in 1995, and then rapidly to 36% in 2002.

Since we are only interested in the effect of college education on earnings, we restrict our sample to individuals who were employed in the survey year with positive earnings. Moreover, since we take into account individuals' self-selection based on heterogeneous returns, we need to assure that individuals in our sample had a reasonable chance of attending college. Before 1977, China's higher education system was severely affected by the Cultural Revolution (1966-1976); many youths were sent to the countryside for "rectification" (or "re-education"), and many colleges and even middle schools were either closed or dysfunctional. In 1977, the government reinstalled the college entrance exams after a ten-year hiatus. After 1978, all high school graduates had reasonable chance of going to college. As a general rule in the late 1970s, children started elementary school at age 7 and stayed for 5 years; junior high school and senior high school each took 2 years. Thus, an individual who was born in 1962 and started

[21] The CHIP-88 and CHIP-95 data are available to the public at the Inter-university Consortium for Political and Social Research (ICPSR). Both data sets have been used intensively by researchers around the world. For details about CHIP-88 and CHIP-95, see Griffin and Zhao (1993) and Riskin, Zhao and Li (2001). CHIP-02 has not yet been released to the public. A recent publication using CHIP-02 is Khan and Riskin (2005).

elementary school at age 7 would be a senior in high school in 1978 and could choose to take the required examinations and go to college. Therefore, we limit all of our samples to individuals born after 1961 in order to avoid abnormalities caused by the Cultural Revolution. Because the official definition of labor force in China starts at age 16, the upper birth-year cutoff point eliminates those born too late to have completed college education by the time of the survey.

The sample is further limited due to availability of family background information such as parental education and income. Our samples are restricted to working individuals who are living in a household with their parents and who have positive earnings in the survey year. As specified in the model, we only include two education groups: 3 or 4-year college graduates and high school graduates.[22] Variable definitions and sample statistics are presented in tables 1, 2a, 2b, and 2c.[23]

In our sample earnings include regular wages, bonuses, overtime wages, in-kind wages, and other income from the work unit. Hourly wage rate is calculated based on the reported number of hours worked. The nominal average hourly wage almost doubled from 2.30 yuan in 1995 to 4.57 yuan in 2002 (with negligible inflation), and the increase is larger for college graduates than for high school graduates. The standard deviation of wage rates also doubled, reflecting a higher degree of wage dispersion. We use parental income as one of the proxies for potential financial constraints on attending college. Ideally we should use parental income at the time when the individual makes the decision (usually the senior year of high school), but that information is not available. We use parental earnings five years prior to the survey date as a proxy for the ideal variable, which is the earliest available earnings information in the data.[24]

---

[22] The education measure includes several degree categories: elementary school or below, junior high, senior high, technical school, junior college (3-year college), and college/university or above. For more details, see Li (2003). Because technical school is different from senior high school and college, we excluded it from our samples. Thus, our samples focus on high school graduates and college graduates.
[23] The sample of 1988 is the largest because we cannot distinguish children and children-in-law in a household. This may cause some problems of mis-matching parents' education and income in the estimation. Yet, this problem should not be very serious for 1988 because the oldest age should be 26 years old, still somewhat too young to be married. In CHIP-95 and CHIP-02, the data can distinguish children and children-in-law.
[24] Such information is not available in CHIP-88, so current parental income is used instead.

The proportion of college graduates in the samples increased from 19% in 1988, to 49% in 1995, and then to 61% in 2002. The rapid increase reflects rapid expansion of the number and capacity of higher education institutions. Tuition also increased over this period. In 1997, the average tuition per student was about 31% of per capita GDP.[25] This ratio reached 53% in 2002.[26]

IV. Empirical Results

We first examine propensity scores of attending college, and then analyze estimates of various treatment effects.

A. Propensity to Acquire a College Education

Table 3a presents simple probit estimates of college attendance in the three sample years, 1988, 1995, and 2002 respectively. The regressors can be roughly categorized into variables related to budget constraint and those related to ability formation. For example, parental income provides the financial resources to attend college. Since individuals in the sample are currently employed, the time they chose to enter college is at least four years prior to the date of the survey. In CHIP-95 and CHIP-02, we have information on parental income up to five years prior to the survey date, and we use parental income five years before the survey as a proxy for any financial constraint affecting college attainment. We include both the incomes of father's and mother's because they may have different effects on education decisions We also include the number of siblings in the household as a proxy for a financial constraint, as children are likely to compete for financial resources to fund education.[27]

Among available variables related to ability formation, parental education is generally viewed as an important factor that contributes to children's ability, both

---

[25] The tuition and enrollment data only include regular institutes of higher education.
[26] According to regulations of the Chinese government in 1999, the maximum tuition shall not exceed 25% of the per student cost of higher education.
[27] Unfortunately, it is an imperfect measure of household size, as not all children lived in the household during the time of survey.

through genetic influences and through parental attitudes toward investment in their children's human capital. We cannot rule out, though, that the financial ability to provide childhood investments in human capital may affect measured ability at older ages (Heckman and Li, 2004), as it reflects past expenditures on learning and nutrition (if income is serially correlated). Other control variables include a gender dummy, an ethnic minority dummy, and a birth year included to capture any year-specific factors related to the opportunities of going to college.[28] One such year-specific factor is related to supply-side constraints on college attendance.

Besides parameter estimates, table 3a also reports the mean marginal propensities (probabilities) attributable to each independent variable. In all three samples, both parental income and education exert positive impact on children's chances of attending college, and in most cases the estimates are also statistically significant. This joint significance implies that parental income and education play distinct roles in children's education attainment despite the high correlation between them. For all three sample years, father's education has a larger effect than that of mother's. The largest difference is found in 1988, but the difference becomes much smaller and negligible in 1995 and 2002.[29]

Mother's income shows a much larger effect on college choice than that of father's in 1988 and 1995, but not in 2002. The marginal effect of mother's income is about four times larger than that of father's in 1988 and two times larger in 1995. The convergence of the marginal effects of father's and mother's income may be explained by parental income serving as a better proxy for ability rather than financial constraints in the earlier period, when tuition was not charged for most colleges and family financial resource wasn't a barrier to college attendance. We postulate that, given father's income, higher mother's income supports better nutrition that contributes to ability formation in children. However, when financial constraints become a significant barrier to college attendance, the importance of mother's income as a financial resource for tuition payment rises and matches that of father's income.

---

[28] Provincial dummies are not included because the current province refers to the individual's place of work and not necessarily the place of attending high school.
[29] The results for 1988 should be interpreted with caution because of possible mismatch of parental education and income.

In 1995, the marginal impact of an additional year of father's education is equivalent to an increase of 5.7 thousand *yuan* of father's income. This value drops to 2.6 thousand *yuan* in 2002, implying a rise in the importance of parental income relative to parental education. We attribute this change to tuition hikes since the mid 1990s. The pattern is not as pronounced for mother's income, though.

The estimated effect of another proxy for family financial constraint — the number of children in the household — is large and significantly negative in 1988. One more sibling reduces the probability of attending college by 4.5%. In 1995 the impact is still a negative 3.7% (and almost statistically significant the 10% level), while in 2002 the impact becomes insignificant. Although this decline in the marginal impact of an additional child would appear to contradict our hypothesis that the influence of financial constrain on college attendance increased over time, we believe that it is due to increasingly stringent enforcement of the one-child policy which substantially reduced variation in the number of children among urban households. The ethnic minority dummy does not appear to be very important in college choice either, although it became quite negative and almost significant at the 10% level in 2002.

Although it may at first appear to be surprising that in 1995 and 2002, the coefficient on the gender (male) dummy is negative and statistically significant at the 10% level, we interpret this higher likelihood for females to attend college as the result of selectivity prior to high-school attendance. In all three samples, female students comprise a smaller proportion of high-school graduates than male students. As far as female students are concerned, enrolling in high school signals strong commitment to attempting college; female students who have completed senior high school are more likely to continue into college.

In table 3a we compare marginal coefficients across years using sample means for each year. In table 3b we perform the same exercise using the overall sample means, i.e. the three-year average. In order to anchor the birth year dummy, we choose the cohort born in 1968 which appears in all three samples.[30] The marginal effects for parental income are calculated based on real income adjusted by the urban CPI.

---

[30] For 1988, those born in 1968 or before were combined into the same cohort.

Table 3b shows that, for one thousand *yuan* increase in the real value of father's income, the probability of going to college increases by 0.5 percentage point in1988 and 4 percentage points in 2002. In 2002, the effect of father's income surpasses education and becomes the dominant factor in college entrance. The impact of mother's income also increases, but by a smaller amount. The growing importance of father's income in college entrance, controlling for parental education, is consistent with the rising impact of higher college costs. The marginal effect of parental education increased sharply from 1988 to 1995, and then declined moderately in 2002. An additional year of father's education improved the probability of going to college by 1 percentage point in 1988, 3.4 percentage points in 1995, and 2.4 percentage points in 2002. The effect of mother's education displays the same pattern but the drop in 2002 is smaller.

These probit models generate a propensity score for each observation, which is the predicted probability of college attendance. The frequency distributions of these propensities show a reduced-form picture of growing college attendance in China (Figure 3).[31] For each year the left panel shows the distribution for all observations (S=1 and S=0), while the right panel shows separate distributions for college graduates and high school graduates. The rightward shift of the combined distributions reflects increasing college enrollment and is consistent with the nearly 80% growth of the proportion of the urban population with at least a college education between 1988 and 1995 and more than 100% growth by 1999, as documented in our data and in other studies as well (for example, Zhang and Zhao 2002, table 4).[32] In 1988, the frequency distribution of high school graduates is supported over a range of propensity scores from approximately zero through nearly 0.6;[33] in 1995, it is supported over the range from approximately zero through 0.9, and by 2002, it is supported over almost the entire range of propensities approaching 1.0. The frequency distribution of college graduates is supported over the range of propensities between approximately zero and 0.7 in 1988, between approximately zero and greater than 0.9 in 1995, and from about 0.1 through 1.0 in 2002.

---

[31] The sample densities are smoothed with Gaussian kernels with optimal bandwidths defined in Silverman (1986).

[32] Note that our samples are based on urban residents, which have much better chances of attending college.

[33] A small support implies that omitted factors play important roles or large unobservable heterogeneity exists in schooling decision.

Examining these distributions more carefully reveals some interesting trends. Table 4 shows that in 1988, 19.2% of the sample were college graduates and had a propensity score equal to or greater than 0.323. We define this propensity score to be the cutoff score. In 1988 10.5% of the entire sample had scores higher than this value yet they didn't go to college, and by construction the same fraction of the entire sample had scores lower than this value yet they did go to college. This ratio rises to 15.9% in 1995 and 15.3% in 2002. Moreover, the percentage of such "misfits" in the unschooled group has increased from 12.9% in 1988, to 27.2% in 1995 and 39.7% in 2002, while in the schooled group this percentage has decreased from 54.6% in 1988, to 38.3% in 1988 and 24.9% in 2002. These patterns suggest that unobserved heterogeneity increased dramatically over our sample period, mostly between 1988 and 1995, and such increased heterogeneity is apparently affecting the unschooled group more than the schooled group. The increased heterogeneity could reflect (1) a growing proportion of agents with unobserved financial constraints and high propensity scores who cannot realize their high potential because they are unable to finance college education; or (2) a growing importance of unobserved comparative advantage. If (1) dominates, then we should observe selection bias and sorting gain diminishing over time for the schooled, and increasing over time for the unschooled; however, if (2) dominates, sorting gains for both groups should increase.

B. College Education and Earnings

Table 5 contains the results of OLS, IV, and semi-parametric local instrumental estimation of the effect of college attendance on earnings. For each sample we present four specifications of the earnings equation. The benchmark specification doesn't include any ability proxy, and in the other three specifications we use parental education, parental income, and both as ability proxies, respectively. The benchmark OLS estimates are commensurate with those reported elsewhere for comparable time periods. They exhibit an upward trend in returns to college education during our sample period.[34] The IV

[34] See Fleisher and Wang (2004) and Li (2003) for estimates and a summary of other studies for the same period.

estimates of the return to college education (all of which use the propensity score as the instrument for college attendance) are similar to the OLS estimates in 1988 but become considerably higher than the OLS estimates in 1995 and 2002. Since in general neither OLS nor conventional IV method consistently estimates the average treatment effect, such difference between OLS and IV does not have clear implications (Caneiro, Heckman, and Vytlacil, 2000). In fact, both OLS and IV underestimate the ATE in all three samples, and the OLS bias increased from a negligible -2.4% in 1988, to -42.6% in 1995 (significant at 5%) and -93.7% in 2002 (significant at 1%).

We now turn to our estimates of returns to schooling based on the semi-parametric local IV estimation. The distinctive feature of this procedure is to estimate the marginal treatment effect of choosing college at set levels of unobserved heterogeneity.[35] Figure 4 depicts the estimated MTE of college education from four specifications of the earnings equation for the years 1988, 1995, and 2002. Inclusion of an ability proxy in the local polynomial regressions usually doesn't change the shape of the MTE, but it does shift the curve upward or downward. An important question is whether either parental income or parental education serves as a legitimate proxy for ability. In estimating the linear part of the earnings equation we find in most cases parental education doesn't affect children's earnings, while parental income, particularly father's income, affects children's earnings in a statistically significant manner.[36] Therefore we have more confidence in the specification where parental income is used as ability proxy, but we also report results from other specifications as robustness tests.

We consistently find that between 1988 and 2002 the average treatment effect — the return to education for a randomly selected individual — has increased substantially. For example, in the specification with parental income as ability proxy, the rate of return for four years of college has increased from an insignificant 6.3% in 1988,[37] to a significant 68.7% in 1995, and then to an even more significant 130.7% in 2002. However, when this dramatic change is decomposed into treatment on the treated (TT)

[35] I.e. set μ_s equal to 0.01, 0.02, 0.03…, and 0.99.
[36] Results from estimating the linear part of the earnings equation (i.e., the double residual regressions) are available on request.
[37] 6.3% is computed with the formula 100[exp(0.061)-1) where 0.061 is taken directly from the corresponding entry in Table 5. All estimates discussed in this paragraph are transformed this way.

and treatment on the untreated (TUT), we obtain strikingly different results. We regard the TT as *realized* return that is achieved by individuals who actually completed four years of college, while TUT as *unrealized* potential return that could have been earned by those who did go to college. This realized return to college has increased from 51.7% in 1988, to 84.0% in 1995, and to 145.0% in 2002. Meanwhile, the unrealized return has also increased from -2.4% to 58.6%, and then to 109.6%. The increase in the unrealized return outpaces that in the realized return. This implies that although the return to those who go to college has increased drastically since 1988, the potential return for those who do not (or cannot) go to college has increased even more.

Further insights are provided by analyzing the changes in selection bias and sorting gain for the schooled and unschooled groups over time. For those who go to college, the selection bias, the mean difference in unobservables between the counterfactual of what a college graduate would earn if he didn't attend college and what an average high school graduate earns, has become more and more negative (-0.222, -0.376, and -0.639 respectively). The sorting gain for college graduates, the counterfactual difference between what an average college graduate earns and what he would earn if the college degree were not obtained, has dwindled (0.356, 0.087, and 0.060 respectively). For those who do not go to college, the selection bias, the mean difference in unobservables between the counterfactual of what a high school graduate would earn had he completed college and what an average college graduate earns has changed from negative to positive and become larger in absolute value (-0.219, 0.227, and 0.483 respectively). The sorting gain for the unschooled group, the difference between what an average high school graduate earns and the counterfactual of what would be earned had he completed college has been small and hardly significant (0.085, 0.062, and 0.096 respectively). There has been a decline in selective sorting according to comparative advantage over time. These findings indicate that the schooled group seems to be able to self-select based on what they do better, that their average gain from attending college has declined over time, and that their relative disadvantage in jobs taken by the unschooled has increased. In contrast, members of the unschooled group apparently have had increasing difficulty choosing their desired education, in that the average unschooled person would gain substantially by becoming schooled.

The estimation results for 1988 are consistent with the implications of the classic model of sorting by comparative advantage: both selection biases are negative while both sorting gains are positive. In 1995 and 2002, the estimated selection biases for the schooled group remain negative, but those for the unschooled group become positive. This is consistent with high school graduates' having higher ability than college graduates. But this seems counterintuitive. An alternative, and we believe more appealing, interpretation is that unobserved barriers to acquiring a higher education explain this phenomenon. It is that college-worthy students are forced into the unschooled group by inability to finance college, and the high-school group has come to include high potential return individuals who would do better than an average college graduate had this individual been able to go to college. We find evidence for this inefficient selection pattern in the 1995 and 2002 samples.

The heterogeneous return model postulates that those who attend college do so because they benefit more than those who choose not to attend. It is important to emphasize that this assumption does imply that decisions are made strictly in terms of expected income streams. It is consistent with someone choosing not to attend college because financial or psychic costs are expected to outweigh financial gains (Carneiro, Heckman, and Vytlacil 2003). However, if all financial and psychic costs of college attendance are reflected in the propensity score, the model implies the MTE function is monotonically negatively sloped and represents a demand for college education in the sense that a decline in the marginal financial cost of college attendance is required to induce greater college attendance, *cet. par.* MTE captures observed gross financial gains from attending college, and it is only in this sense we identify those with high MTE as most "college-worthy". On the other hand, the unobserved heterogeneity in schooling decision, μ_s, increases from left to right in Figure 4 and indicates the (declining) probability of actual attendance of college. The MTE curves for 1988 support this hypothesis: people with high gross financial returns are also more likely to attend college. However, the MTE curves become U-shaped in 1995 and 2002. This shape implies that some of the highly college-worthy individuals have difficulty attending college. These shapes are inconsistent with the joint hypothesis that agents' unobserved heterogeneity involves only their comparative advantage in ability to benefit from more schooling.

They are consistent with unobserved barrier to college attendance in China, e.g. psychic costs or unobserved financial barriers (Carneiro, Heckman, and Vytlacil 2004, p. 25).

Finally, we briefly discuss the implications of differences across the four specifications. As described in Section II, variables included in the schooling choice (Z) that are excluded from the earnings equation (X) serve as identifying instruments. For example, the instruments used in the benchmark model include parental income and education, number of siblings, and birth year. In the specification where parental income is used as an ability proxy, this variable is included in the earnings equation and is excluded from the list of instruments. So the question boils down to whether parental income or education affects children's income *beyond* their impact through college education. If parental income and/or parental education are important determinants of children's earnings, omitting them in the earnings equation potentially biases the estimates. On the other hand, if it turns out that neither of them matters in the earnings equation, including them reduces the number of valid instruments and consequently the efficiency of estimation. That is, as we add more variables to the right-hand-side of the earnings equation, the estimates are less likely to be biased, but more likely to be inefficient. While we cannot establish which of the MTE estimates illustrated in figure 4 is based on the least biased or most efficient estimates of the earnings equation, the change in MTE over time is strikingly consistent. Whereas in the 1988 data, the slope of the MTE curve is uniformly negative, indicating efficient, purposive selection into college education, the 1995 and 2002 curves are persistently U-shaped, indicating that among the unschooled there is a large proportion of individuals who would gain more from college attendance than those who have chosen to extend their schooling beyond high school. We believe this result has important policy implications.

V. Conclusion

All three estimation methods — OLS, IV, and semi-parametric local IV (SPIV) — revealed substantial increase in returns to schooling in China between 1988 and 2002. However, they differ substantially in the estimated levels of returns. Only the SPIV estimates distinguish different groups and answer well-posed policy questions. We find that the increase in the average treatment effect is driven by both an increase in the returns to those who have chosen college education, but also by a large, if not larger increase in the potential returns to those who remain in the "unschooled" state. Additional evidence of some kind of mis-sorting is that our estimates on selection bias indicate that while purposive selection has increased for college graduates, it has declined for high school graduates. In addition, sorting gain has become less pronounced for college graduates but remained small for high school graduates. We interpret these results as evidence indicating that the higher education system has become more efficient in terms of rewarding the schooled group, but less efficient in terms of failing to allow a growing number of college-worthy youth into college.

Our sample covers a period in which China's higher education system underwent major structural changes. Higher costs of college education affect self-selection in two ways. Individuals (and their families) in the schooled group have responded to higher expected returns and have willingly paid the higher costs of choosing college. On the other hand, among the unschooled group, we find evidence that individuals who would reap a return more than sufficient to compensate for the costs of college attendance (as evidenced by the choices of those who attended college) have chosen not to go to college. This suggests to us that either the distaste for college education has increased over time, or that financial constraints on college attendance have become more severe. We find the first explanation implausible and the second one likely in light of the changes in education finance in China. If our interpretation is correct, the movement toward higher tuition and private funding of higher education, while justified on many grounds, will also contribute to increasing income inequality in a vicious cycle. More specifically, those from wealthy families are more likely to reap the higher returns of education and thus will become wealthier; however, those from poor families may be excluded from schooling opportunities and thus remain poor. Therefore, government policies that help individuals from financially-disadvantaged families gain access to higher education have become crucial and are imperative to help everyone gain equal opportunity to the benefits of education reforms in China.

Labor market reform is a critical component of the transition from planning to markets, and the old wage-grid and *nomenklatura* systems are bound to be replaced by a market system that reflects the true value of education and allows individuals to make the schooling choices that they deem to be optimal. The main policy implication of our results is that labor market and schooling reform without capital market reform and some kind of safety net for the financially disadvantaged robs economic transition of many of its potential benefits.

We would have preferred to have available some kind of standard ability measure in order to obtain estimates of our earnings equations that do not require the proxy measures we have been forced to use. This defect in our data notwithstanding, we find that the pattern over time of our empirical results is quite robust to alternative specifications, and this greatly increases our confidence in the interpretation that mis-sorting of individuals in their schooling choices has increased under the new educational funding system in China.

.

References

Björklund, Anders, and James Moffitt, “The Estimation of Wage Gains and Welfare Gains in Self-Selection Models,” *Review of Economics and Statistics* 69:1 (1987), 42-49.

Brainerd, Elizabeth, “Winners and Losers in Russia’s Economic Transition,” *American Economics Review* 88:5 (December 1998), 1094-1116.

Carneiro, Pedro, and James J. Heckman, “The Evidence on Credit Constraints in Post-Secondary Schooling,” *Economic Journal* 112 (October 2002), 705-734.

Carneiro, Pedro, James J. Heckman, and Edward Vytlacil, “Estimating the Returns to Education when It Varies Among Individuals,” University of Chicago working paper (2000).

Carneiro, Pedro, James J. Heckman, and Edward Vytlacil, “Understanding What Instrumental Variables Estimate: Estimating the Average and Marginal Returns to Schooling,” University of Chicago working paper (2003).

Fan, J., “Design-adaptive Nonparametric Regression,” *Journal of the American Statistical Association* 87 (1992), 998-1004.

Fan, J., “Local Linear Regression Smoothers and Their Minimax Efficiencies,” *The Annals of Statistics* 21 (1993), 196-216.

Fan, J., and I. Gijbels, *Local Polynomial Modeling and Its Applications* (Chapman & Hall, 1996).

Fleisher, Belton, “Higher Education in China: A Growth Paradox,” in Yum K. Kwan, and Eden S. H. Yu (Eds.), *Critical Issues in China’s Growth and Development* (Ashgate Publishing Ltd., Aldershot, UK, 2005), 3-21.

Fleisher, Belton, and Jian Chen, “The Coast-Noncoast Income Gap, Productivity, and Regional Economic Policy in China," *Journal of Comparative Economics* 25:2 (1997), 220-36.

Fleisher, Belton, Keyong Dong, and Yunhua Liu, “Education, Enterprise Organization, and Productivity in the Chinese Paper Industry," *Economic Development and Cultural Change* 44:3 (April 1996), 571-587.

Fleisher, Belton, Klara Sabirianova, and Xiaojun Wang, "Returns to Skills and the Speed of Reforms: Evidence from Central and Eastern Europe, China, and Russia," *Journal of Comparative Economics* 33:2 (June 2005), 351-370.

Fleisher, Belton, and Xiaojun Wang, "Skill Differentials, Return to Schooling, and Market Segmentation in a Transition Economy: The Case of Mainland China," *Journal of Development Economics* 73:1 (2004), 715-728.

Fleisher, Belton, and Xiaojun Wang, "Returns to Schooling in China under Planning and Reform," *Journal of Comparative Economics* 33:2 (2005), 265-277.

Fleisher, Belton, and Dennis Tao Yang, "China's Labor Markets," Ohio State University working paper (2003).

Giles, John, Albert Park, and Juwei Zhang, "The Great Proletarian Cultural Revolution, Disruption to Education, and Returns to Schooling," Michigan State University working paper (2004).

Griffin, K., and R. Zhao (Eds.), *The Distribution of Income in China* (New York, St. Martin's Press, 1993).

Griliches, Zvi,"Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica* 45:1 (January 1977), 1-22.

Hannum, Emily, and Meiyan Wang, "Geography and Educational Inequality in China," University of Pennsylvania working paper (2004).

Heckman, James J., "China's Human Capital Investment," *China Economic Review* 16:1 (2005), 50-70.

Heckman, James, Hidehiko Ichimura, Petra Todd, and Jeffrey Smith, "Characterizing Selection Bias Using Experimental Data," *Econometrica* 66:5 (1998), 1017-1098.

Heckman, James J., and Xuesong Li, "Selection Bias, Comparative Advantage, and Heterogeneous Returns to Education: Evidence from China in 2000," *Pacific Economic Review* 9 (2004), 155-171.

Heckman, James J., and Edward Vytlacil, "Local Instrumental Variable and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences* 96 (1999), 4730-4734.

Heckman, James J., and Edward Vytlacil, "Local Instrumental Variables," in C. Hsian, K. Morimune, and J. Powells (Eds.), *Nonlinear Statistical Modeling: Proceedings of* *the Thirteenth International Symposium in Eco0nomic Theory and Econometrics: Essays in Honor of Takeshi Amemiya* (Cambridge: Cambridge Unviersity Press, 2000), 1-46.

Ichimura, Hidehiko, and Petra E. Todd, “Implementing Nonparametric and Semiparametric Estimators,” manuscript prepared for *Handbook of Econometrics* 5, 2004.

Jones, Derek C., and K. Ilayperuma, “Wage Determination under Plan and Early Transition: Bulgarian Evidence using Matched Employer-Employee Data,” *Journal of Comparative Economics* 33:2 (June 2005), 227-243.

Khan, Azizur R., and Carl Riskin, “China's Household Income and Its Distribution, 1995 and 2002,” *The China Quarterly* 182 (2005), 356-384.

Knight, John, and Lina Song, “The Determinants of Urban Income Inequality in China,” *Oxford Bulletin of Economics and Statistics* 53:2 (1991), 123-154.

Li, Haizheng, “Economic Transition and Returns to Education in China,” *Economics of Education Review* 2 (2003), 317-328.

Liu, Zhiqiang, “Earnings, Education, and Economic Reform in China,” *Economic Development and Cultural Change* 46:4 (1998), 697-726.

Meng, Xin, *Labour Market Reform in China* (Cambridge, UK: Cambridge University Press, 2000).

Meng, Xin, and R. G. Gregory, “The Impact of Interrupted Education on Subsequent Educational Attainment: A Cost of the Chinese Cultural Revolution,” *Economic Development and Cultural Change* 50:4 (2002), 934-955.

Meng, Xing, and Michael P. Kidd, “Labor Market Reform and the Changing Structure of Wage Determination in China’s State Sector during the 1980s,” *Journal of Comparative Economics* 25:3 (1997), 403-421.

Munich, Daniel, Jan Svejnar, and Katherine Terrell, “Returns to Human Capital under the Communist Wage Grid and During the Transition to a Market Economy,” *Review of Economics and Statistics* 87:1 (2000), 100-123.

Orazem, Peter, and Milan Vodopivec, “Winners and losers in the Transition: Returns to Education, Experience, and Gender in Slovenia,” *The World Bank Economic Review* 9:2 (1995), 201-230.

Riskin, C., R. Zhao, and S. Li (Eds.), *China's Retreat from Equality: Income Distribution and Economic Transition* (Armonk, N.Y., M.E. Sharpe, 2001).

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

Silverman, B. W., *Density Estimation for Statistics and Data Analysis* (Chapman & Hall, 1986).

Wang, Sangui, Dwayne Benjamin, Loren Brandt, John Giles, Yingxing Li, and Yun Li, "Inequality and Poverty in China during Reform," PMMA working paper (2007).

Willis, R., "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions," in O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics* (Amsterdam: North-Holland, 1986).

Willis, R., and S. Rosen, "Education and Self-Selection," *Journal of Political Economy* 87:5 (part 2, 1979), S7-36.

Yang, Dennis Tao, "Urban-Biased Policies and Rising Income Inequality in China," *American Economic Review* 89:2 (1999), 306-310.

Yang, Dennis Tao, "What Has Caused Regional Inequality in China?" *China Economic Review* 13:4 (2002), 331-334.

Yang, Dennis Tao, "Determinants of Schooling Returns during Transition: Evidence from Chinese Cities," *Journal of Comparative Economics* 33:2 (2005), 244-264.

Zhang, Junsen, and Yaohui Zhao, "Economic Returns to Schooling in Urban China, 1988-1999," paper presented at the meetings of the Allied Social Sciences Association, Washington DC, 2002.

Zhou, Xueguang, "Economic Transformation and Income Inequality in Urban China: Evidence from Panel Data," *American Journal of Sociology* 105:4 (January 2000), 1135-74.

Zhou, Xueguang, and Liren Hou, "Children of the Cultural Revolution: The State and the Life Course in the People's Republic of China," *American Sociological Review* 65:1 (February 1999), 12-36.

Zhou, Xueguang, and Phyllis Moen, *The State and Life Chances in urban China, 1949-1994* (Computer file), ICPSR version. Durham, NC: Duke University, Dept of

Sociology [producer]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. 2002.

Table 1.—Variable Definition

Variable	1988	1995	2002
Father Education (years)	Father's Education		
Mother Education (years)	Mother's Education		
Father's Salary in Earlier Years (1000 Yuan)	Father's Annual Salary in 1988	Father's Annual Salary in 1990	Father's Annual Salary in 1998
Mother's Salary in Earlier Years (1000 Yuan)	Mother's Annual Salary in 1988	Mother's Annual Salary in 1990	Father's Annual Salary in 1998
Number of Children	Number of Children living in the household		
Gender (male=1)	Gender dummy		
Ethnic	1 if the individual is an ethnic minority (non-Han Chinese)		
Work Experience (years)	Estimated by age minus years of schooling minus 6	Year of Work Experience Reported	
Wage	Monthly Wage	Hourly Wage Rate	
Government Sector	Government or public institutions	Not available	
State-owned Sector	State-owned at central or provincial governmental level		
Local Publicly-owned Sector	Publicly-owned at lower government level		
Urban Collective Sector	Collectively Owned Sector		
Province	Dummy variables for each province		
Industry	Dummy variables for each industry		
Birth Year	Dummy variables for the year of birth		
College	1 if individual is a college graduate		

Note: wage includes regular wage, bonus, subsidies and other income from the work unit.

Table 2a.—Descriptive Statistics for 1988 Sample

Variable	Full Sample		College Graduates		High School Graduates	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Father Education (years)	10.19	3.40	11.71	3.50	9.83	3.28
Mother Education (years)	8.33	3.45	9.39	3.83	8.08	3.31
Father Salary 1988 (1000 Yuan)	2.12	1.31	2.30	1.56	2.08	1.24
Mother Salary 1988 (1000 Yuan)	1.22	1.06	1.39	1.08	1.19	1.06
Number of Children	1.83	0.85	1.64	0.73	1.88	0.87
Gender (Male=1)	0.50	0.50	0.51	0.50	0.50	0.50
Ethnic	0.03	0.18	0.04	0.19	0.03	0.18
Monthly Salary (1000 Yuan)	0.15	0.48	0.26	0.44	0.13	0.49
Work Experience (years)	4.26	2.35	2.71	1.68	4.63	2.34
State-owned Sector	0.42	0.49	0.52	0.50	0.40	0.49
Local Publicly-owned Sector	0.40	0.49	0.42	0.49	0.40	0.49
Urban Collective Sector	0.17	0.37	0.06	0.24	0.19	0.39
College	0.19	0.39	-	-	-	-
Number of Observations	1128		216		912	

Note: the omitted ownership sector is non-public sector including private enterprises, Sino-foreign joint venture, etc.

Table 2b.—Descriptive Statistics for 1995 Sample

Variable	Full Sample		College Graduates		High School Graduates	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Father Education (years)	11.70	3.27	12.89	2.97	10.86	3.21
Mother Education (years)	9.80	3.50	11.04	3.35	8.91	3.33
Father Salary 1990 (1000 Yuan)	3.55	2.13	3.76	2.49	3.41	1.83
Mother Salary 1990 (1000 Yuan)	2.58	1.55	2.87	1.77	2.38	1.34
Number of Children	1.66	0.65	1.61	0.64	1.70	0.65
Gender (male=1)	0.60	0.49	0.59	0.49	0.62	0.49
Ethnic	0.04	0.19	0.04	0.19	0.03	0.18
Hourly Wage (Yuan/hour)	2.30	1.84	2.69	2.21	2.02	1.46
Work Experience (years)	5.55	3.77	5.02	3.55	5.94	3.87
State-owned Sector	0.26	0.44	0.32	0.47	0.21	0.41
Local Publicly-owned Sector	0.55	0.50	0.52	0.50	0.56	0.50
Urban Collective Sector	0.10	0.31	0.06	0.24	0.13	0.34
College	0.42	0.49	-	-	-	-
Number of Observations	686		285		401	

Note: the omitted ownership sector is non-public sector including private enterprises, Sino-foreign joint venture, etc.

Table 2c.—Descriptive Statistics for 2002 Sample

Variable	Full Sample		College Graduates		High School Graduates	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Father Education (years)	10.57	3.22	11.23	3.25	9.53	2.89
Mother Education (years)	9.54	3.06	10.04	2.94	8.75	3.08
Father Salary 1998 (1000 Yuan)	10.24	6.54	11.17	7.27	8.75	4.82
Mother Salary 1998 (1000 Yuan)	6.97	4.74	7.52	5.29	6.10	3.54
Number of Children	1.26	0.46	1.27	0.47	1.24	0.45
Gender (male=1)	0.61	0.49	0.54	0.50	0.71	0.45
Ethnic	0.05	0.22	0.04	0.20	0.07	0.25
Work Experience (years)	6.46	4.96	5.82	4.54	7.46	5.42
Hourly Wage (Yuan/hour)	4.57	3.64	5.21	4.16	3.56	2.28
Government Sector	0.30	0.46	0.42	0.49	0.11	0.31
State-owned Sector	0.12	0.33	0.10	0.30	0.15	0.36
Local Publicly-owned Sector	0.16	0.37	0.13	0.34	0.22	0.41
Urban Collective Sector	0.06	0.24	0.03	0.18	0.11	0.31
College	0.61	0.49	-	-	-	-
Number of Observations	654		402		252	

Note: the omitted ownership sector is non-public sector including private enterprises, Sino-foreign joint venture, share-holding company, etc.

Table 3a.—Propensity Estimates

	1988 (based on current year parental income)				1995 (based on parental income in 1990)				2002 (based on parental income in 1998)			
Variable	**Param.**	**t-ratio**	**p-value**	**Mean Marginal Effect**	**Param.**	**t-ratio**	**p-value**	**Mean Marginal Effect**	**Param.**	**t-ratio**	**p-value**	**Mean Marginal Effect**
Constant	-2.475	-10.043	0.000		-1. 804	-6.182	0.000		-1.847	-4.588	0.000	
Father Education	0.082	4.849	0.000	**0.020**	0.087	4.543	0.000	**0.034**	0.063	3.288	0.001	**0.023**
Mother Education	0.002	0.132	0.448	**0.001**	0.074	4.030	0.000	**0.029**	0.057	2.784	0.003	**0.021**
Father Salary	0.028	0.794	0.214	**0.007**	0.015	0.533	0.297	**0.006**	0.026	2.230	0.013	**0.010**
Mother Salary	0.098	1.989	0.023	**0.024**	0.032	0.776	0.219	**0.012**	0.021	1.400	0.081	**0.008**
Number of Children	-0.222	-3.664	0.000	**-0.045**	-0.098	-1.201	0.115	**-0.037**	0.023	0.191	0.424	**0.009**
Gender	-0.087	-0.922	0.178	**-0.020**	-0.140	-1.297	0.098	**-0.054**	-0.539	-4.630	0.000	**-0.195**
Ethnic	0.072	0.287	0.387	**0.017**	0.177	0.643	0.260	**0.070**	-0.295	-1.221	0.111	**-0.114**
Number of Observations	1128				686				654			
Log Like.	-466.300				-407.807				-366.503			
Like. Ratio	0.154				0.124				0.159			
Pseudo-R2	0.316				0.510				0.690			

Notes: The dependent variable is binary, which is 1 for graduated from 3- or 4-year college and 0 otherwise. For 1995 and 2002, father's and mother's salary is for the time 5 years prior to the sample year; while for 1988, they are based on the current year income due to data limitation. The results for birth year dummies are not reported. The marginal effects are calculated using (mean+1) for easy interpretation, based on the sample average of each year. For dummy variables, the marginal effects are calculated based on changing its value from 0 to 1.

Table 3b.—Marginal Effect Estimates

	1988	1995	2002
Variable	Mean Marginal Effect	Mean Marginal Effect	Mean Marginal Effect
Father Education	0.010	0.034	0.024
Mother Education	0.0002	0.029	0.021
Father Salary	0.0047	0.022	0.040
Mother Salary	0.020	0.043	0.032
Number of Children	-0.022	-0.037	0.009
Gender	-0.010	-0.054	-0.20
Ethnic	0.009	0.069	-0.12

Notes: The marginal effects are evaluated at the overall sample mean (three years) for each variable. It is calculated as (mean+1) in order to have a more meaningful interpretation. For dummy variables, the marginal effect is calculated based on the difference between 0 and 1. For birth year dummies, we use the 1968 cohort as the default. The marginal effects for father and mother's salary are measured by 1,000 Yuan increase in 1984 value based on the urban CPI index (1988, 157.2; 1995, 358.4; 2002, 396.3).

Table 4.—Comparison of Propensity Distributions

	1988	1995	2002
Sample size	1128	686	654
Number of non-attenders	912	401	252
Number of attenders	216	285	402
Proportion of sample who are college attenders or graduates	19.15%	41.55%	61.47%
Cut-off Propensity	0.323	0.459	0.554
Number of respondents in the "wrong" group	118	109	100
Percentage of the non-attender group	12.94%	27.18%	39.68%
Percentage of the attender group	54.63%	38.25%	24.88%
Percentage of the total sample	10.46%	15.89%	15.29%

Notes: The cut-off percentage is the propensity score that corresponds to the cumulative frequency of the total sample that was attending or had graduated from college in the sample year.

Table 5: Marginal Treatment Effect Estimates

Parameter	CHIP88				CHIP95				CHIP02			
Ability proxy	None	Edu	Income	Both	None	Edu	Income	Both	None	Edu	Income	Both
OLS	0.202[a] (0.033)	0.205[a] (0.039)	0.195[a] (0.042)	0.203[a] (0.030)	0.259[a] (0.049)	0.223[a] (0.056)	0.234[a] (0.051)	0.216[a] (0.056)	0.309[a] (0.053)	0.277[a] (0.057)	0.257[a] (0.054)	0.246[a] (0.053)
IV[*]	0.192[a] (0.066)	0.205[a] (0.072)	0.078 (0.064)	0.135[a] (0.053)	0.548[a] (0.169)	0.262 (0.316)	0.353[b] (0.164)	0.052 (0.299)	0.948[a] (0.146)	0.979[a] (0.238)	0.603[a] (0.169)	0.632[a] (0.182)
ATE	0.225[b] (0.134)	0.363[b] (0.165)	0.061 (0.169)	0.218 (0.207)	0.685[a] (0.280)	0.612[c] (0.395)	0.523[b] (0.266)	0.351 (0.391)	1.246[a] (0.248)	1.467[a] (0.325)	0.836[a] (0.270)	0.975[a] (0.305)
TT	0.592[a] (0.201)	0.873[a] (0.258)	0.417[b] (0.240)	0.721[a] (0.200)	0.815[b] (0.379)	0.721[c] (0.547)	0.610[c] (0.371)	0.396 (0.483)	1.103[a] (0.337)	1.289[a] (0.349)	0.896[a] (0.325)	1.012[a] (0.312)
TUT	0.138 (0.150)	0.241[c] (0.181)	-0.024 (0.202)	0.099 (0.246)	0.591[b] (0.343)	0.535[c] (0.389)	0.461[c] (0.326)	0.318 (0.426)	1.474[a] (0.304)	1.751[a] (0.417)	0.740[b] (0.321)	0.915[b] (0.432)
Bias =OLS-ATE $=E(U_1 \mid S=1)-E(U_0 \mid S=0)$	-0.024 (0.127)	-0.157 (0.150)	0.134 (0.159)	-0.015 (0.196)	-0.426[b] (0.268)	-0.340 (0.387)	-0.289 (0.258)	-0.135 (0.394)	-0.937[a] (0.244)	-1.190[a] (0.318)	-0.579[b] (0.259)	-0.729[a] (0.307)
Selection Bias (S=1) =OLS-TT $=E(U_0 \mid S=1)-E(U_0 \mid S=0)$	-0.390[b] (0.188)	-0.668[a] (0.233)	-0.222 (0.215)	-0.517[a] (0.183)	-0.557[c] (0.368)	-0.498 (0.536)	-0.376 (0.361)	-0.180 (0.483)	-0.794[a] (0.337)	-1.013[a] (0.344)	-0.639[b] (0.316)	-0.766[a] (0.315)
Sorting Gain (S=1) =TT-ATE $=E(U_1-U_0 \mid S=1)$	0.367[b] (0.183)	0.511[b] (0.226)	0.356[c] (0.258)	0.502[b] (0.240)	0.131 (0.264)	0.109 (0.285)	0.087 (0.261)	0.045 (0.266)	-0.143 (0.166)	-0.178 (0.149)	0.060 (0.142)	0.037 (0.158)
Selection Bias (S=0) TUT – OLS $=E(U_1 \mid S=0)-E(U_1 \mid S=1)$	-0.064 (0.153)	0.036 (0.172)	-0.219 (0.200)	-0.104 (0.238)	0.332 (0.334)	0.312 (0.387)	0.227 (0.324)	0.102 (0.446)	1.165[a] (0.296)	1.474[a] (0.409)	0.483[c] (0.314)	0.669[c] (0.433)
Sorting Gain (S=0) ATE – TUT $=E(U_0-U_1 \mid S=0)$	0.087[b] (0.041)	0.122[a] (0.052)	0.085[c] (0.061)	0.119[b] (0.059)	0.094 (0.188)	0.077 (0.205)	0.062 (0.188)	0.033 (0.196)	-0.228 (0.269)	-0.284 (0.240)	0.096 (0.226)	0.060 (0.250)
TT - TUT	0.454[b] (0.214)	0.632[b] (0.277)	0.441[c] (0.319)	0.622[b] (0.298)	0.224 (0.452)	0.187 (0.489)	0.149 (0.449)	0.077 (0.431)	-0.371 (0.434)	-0.462 (0.389)	0.156 (0.368)	0.097 (0.408)

Notes: Dependent variable is monthly wage in 1988, hourly wage in 1995 and 2002. OLS regressors are a binary variable for college attendance, experience, experience squared, a dummy variable = 1 if male, a dummy variable = 1 if ethnicity not Han Chinese. The IV regression uses the propensity score as the instrument. The treatment effect estimates are based on results from local polynomial regression. Standard errors shown in parentheses are obtained by bootstrapping, and a superscript "a" denotes statistical significance level of 0.01, "b" of 0.05, and "c" of 0.10. All coefficients represent the estimated return to four years of college.

Figure 1.—Heterogeneous Returns and Instrumental Variable Method

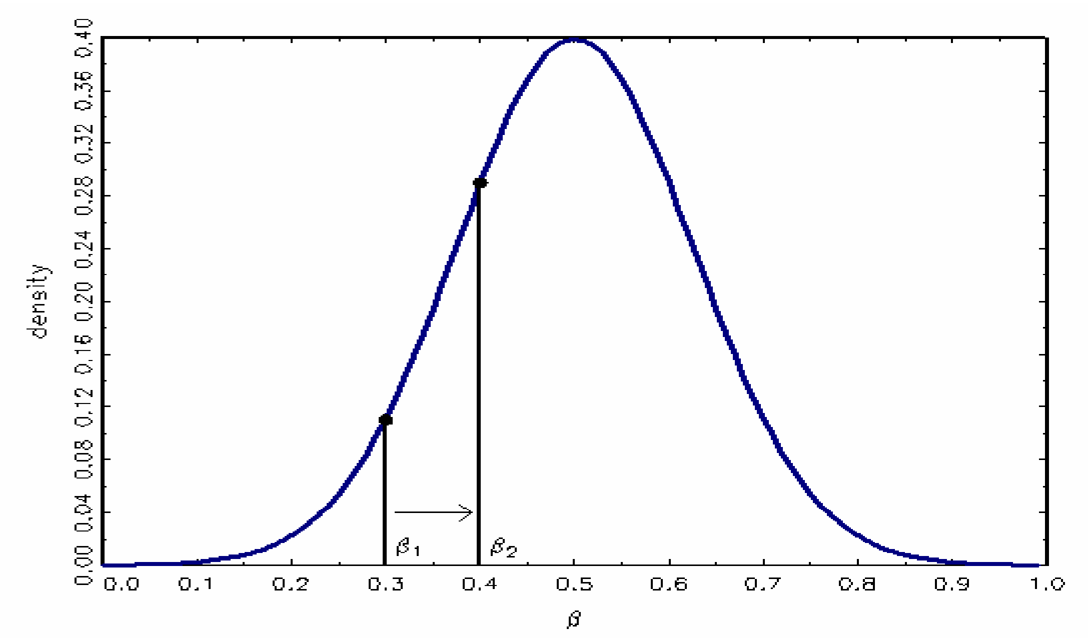

Figure 2.—Estimating the Marginal Treatment Effect – A Visual Exposition

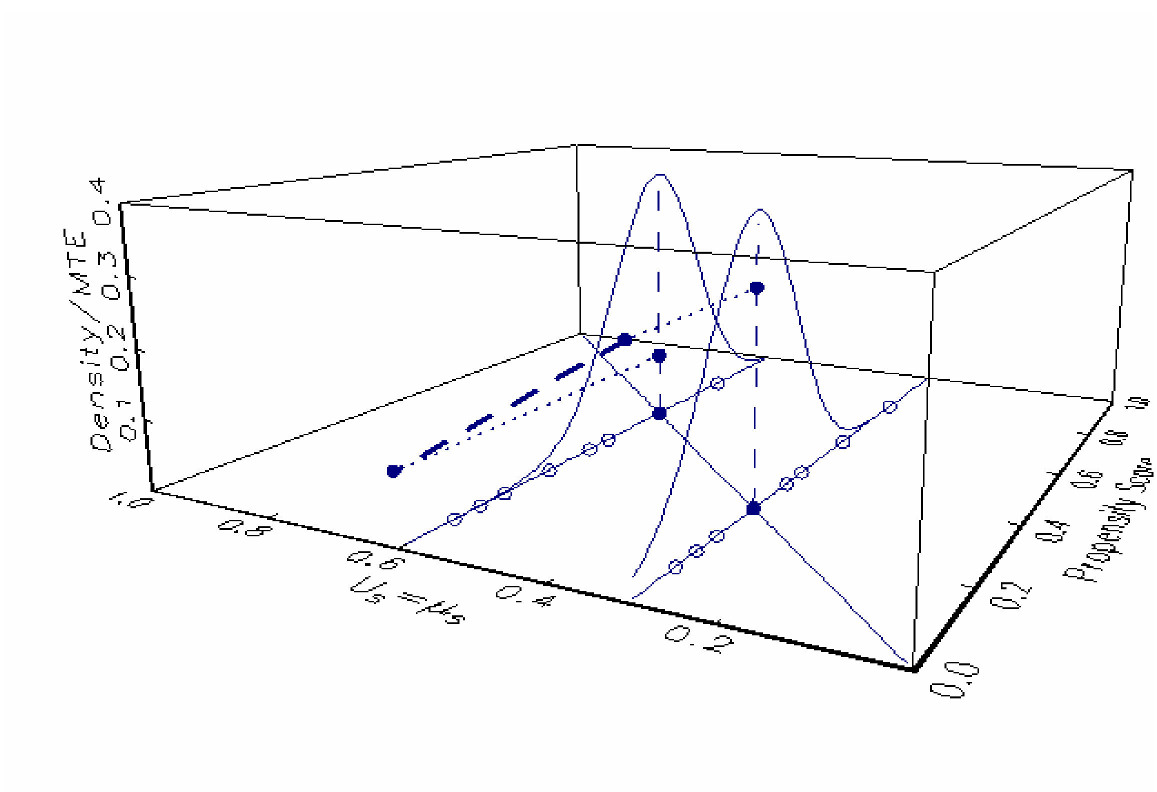

Figure 3.—Propensity Score Distribution: Kernel-Smoothed

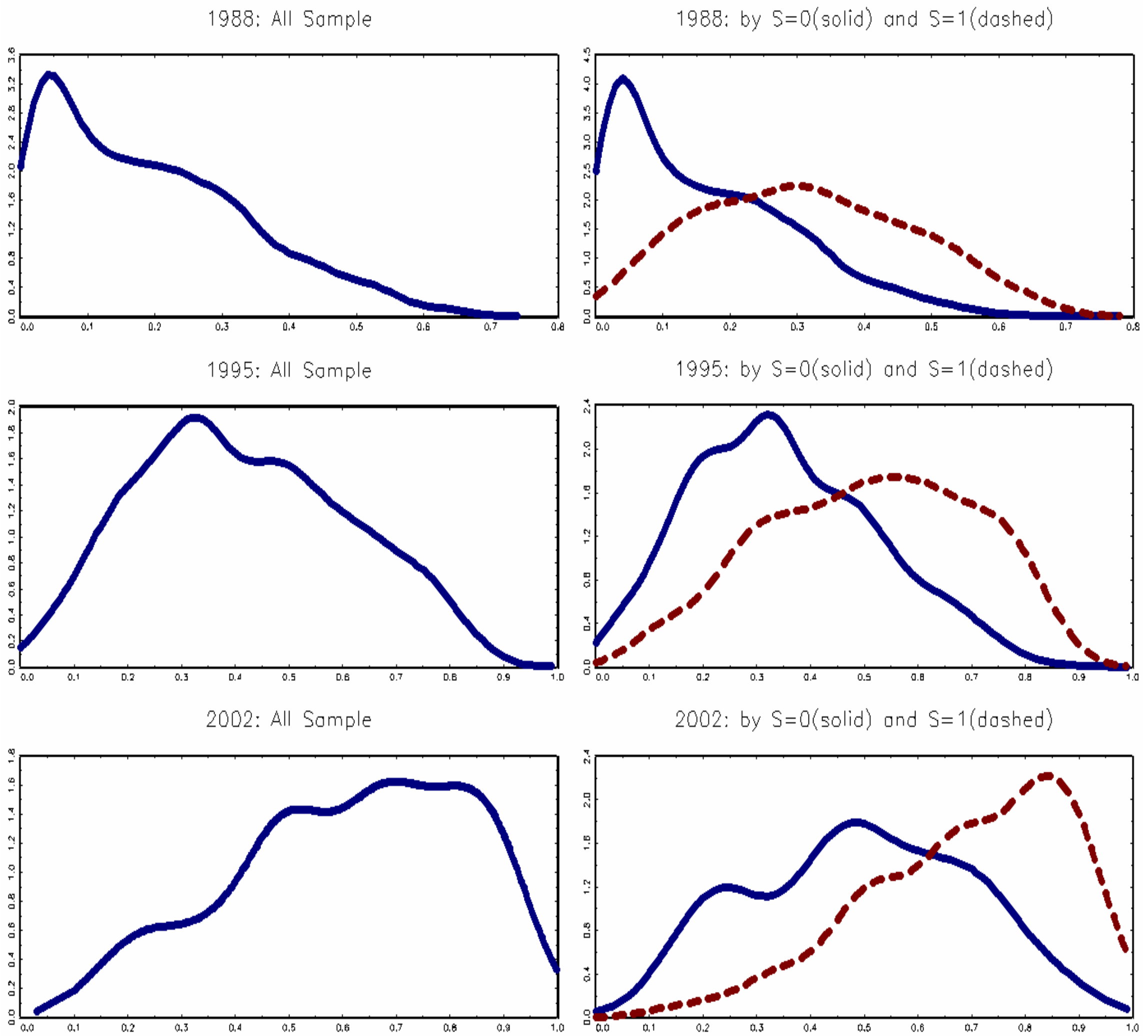

Figure 4.—Marginal Treatment Effects