DISCUSSION PAPER SERIES

IZA

IZA DP No. 1446

Sorting, Selection, and Transformation of the Return to College Education in China

Belton M. Fleisher
Haizheng Li
Shi Li
Xiaojun Wang

December 2004

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Sorting, Selection, and Transformation of the Return to College Education in China

Belton M. Fleisher
Ohio State University

Haizheng Li
Georgia Tech

Shi Li
Chinese Academy of Social Sciences and IZA Bonn

Xiaojun Wang
University of Hawaii at Manoa

Discussion Paper No. 1446
December 2004

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

IZA Discussion Paper No. 1446
December 2004

ABSTRACT

Sorting, Selection, and Transformation of the Return to College Education in China

We estimate selection and sorting effects on the evolution of the private return to schooling for college graduates during China's between 1988 and 2002. We pay special attention to the changing role of sorting by ability versus budget-constraint effects as China's education policy has changed from one in which the bulk of direct costs are paid by government for students who pass a rigid set of test to one in which freedom of choice is increasingly the rule for those who can afford to pay for tuition and living expenses while acquiring higher education. We find evidence of substantial sorting gains under the traditional system but that gains have diminished and even become negative as schooling choices widened and participation has become subject to increasing direct private costs. We take this as evidence consistent with the influence of financial constraints on decisions to attend college.

JEL Classification: J31, J24, O15

Keywords: return to schooling, sorting gains, heterogeneity, financial constraints, comparative advantage

Corresponding author:

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

1. Introduction and Background

From about the inception of economic reform in China into the early 1990s, wage differences by level of skill, occupation, and/or schooling remained very narrow. The Mincerian return to higher education was quite low in comparison with that in the early years of the Mao era and than in other industrialized and industrializing countries including those in some smaller transition economies, such as the Czech Republic, Slovenia, and Bulgaria [2]. Fleisher and Wang (2004c) show that the time path of the return to college education paralleled that to schooling in general. Moreover, college graduates appear to have been severely underpaid relative to their contribution to production (Fleisher and Wang, 2004a). There is evidence that in the past 15 years, returns to schooling in China have begun to increase (Zhang and Zhao, 2002; Li, 2003, Yang, 2004). This movement toward what probably more closely approximates a market-determined rate of return to schooling has paralleled rising income inequality, and while it has probability contributed to this growing income disparity it seems clear that other factors dominate. According to Yang (1999), China in the late 1990a 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 how rising inequality in China is associated with access to educational opportunities. The proportion of the population privileged to attend college has been and remains very extremely small by almost any standard, despite a sharp acceleration of schooling expenditures in the past decade (Fleisher and Wang, 2004b; Heckman, 2004). The proportion of the population aged 20 and higher with a college degree was less than 3.2% in 1993 and grew to 3.5% according to 1993 and 2000 population census (National Bureau of Statistics of China, 1994 and 2002). If access to higher levels of schooling is available only to the 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. Although 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 junior- and senior high schools, more recently, a growing proportion of college students must fund their own educational expenses (Hannum, 2004; Heckman, 2004). Access to college, and an individual's chances of economic gain from college, depend on the ability to achieve high test scores and on cognitive and noncognitive attributes produced in earlier family and educational contexts. These traits, in turn, depend 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). This raises the question whether increased public spending on education in China has enabled those who will most benefit themselves (and society) to achieve higher educational attainment or whether growing reliance on private funding through tuition and other costs have led to adverse sorting based on ability to pay rather than comparative advantage.

In this paper we focus on the returns to college education in China from the end of the first decade of transition to 2002, paying particular attention to sorting and selection issues . We address the following questions in this paper:

1. How have the relative importance variables that determine the probability of college attendance changed?
2. Is there evidence that sorting into college has become more or less efficient during reform?
    - Has the sorting gain narrowed or widened?
    - If it has widened, is this because more able students are now able to attend college due to reduced favoritism, or is it because lower levels of schooling provide worse training for college?

- If it has widened, is there evidence of a growing gap across income categories, suggesting either increased importance of borrowing constraints or long-run income effects on ability to benefit from college (Carneiro and Heckman, 2002)?

2. Methodology

The marginal treatment effect (MTE) and its derivatives are estimated using the method developed in Heckman, Ichimura, Todd, and Smith (1998).[3, 4]

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

$$\ln Y_1 = \mu_1(X, U_1)$$
$$\ln Y_0 = \mu_0(X, U_0)$$

where a subscript indicates whether the individual is in the schooled state or the unschooled state. Y is income, X is observed heterogeneity, and U is unobserved heterogeneity in wage determination. In general, the functional forms can have a nonlinear component, and $U_1 \neq U_0$.

The schooling choice comes from the following latent dependent model:

$$S^* = \mu_s(Z) - U_s$$
$$S = 1 \ \ if \ \ S^* \geq 0$$

where S^* is a latent variable whose value is determined by an observable component $\mu_s(Z)$ and a unobservable component U_S.

In our empirical work, Z is a vector of variables that help predict the probability of attending college. It includes parental education, parental income, number of children, gender, ethnic group, and birth year dummies. X is a vector that holds explanatory power on wages. In the benchmark setting, this includes experience, experience squared, gender, ethnic group, ownership, industry, and location.

In the first step, a probit model is used to estimate the $\mu_s(Z)$ function. The predicted value is called propensity score, $\hat{P}_i$, where the subscript i indicates each individual. The second step adopts a semi-parametric procedure in which *local linear regressions* are used frequently. Fan (1992, 1993)[5] develops the distribution theory for the local linear estimator of $E(Y|P=P_0)$, where Y and P are random variables. $E(Y|P=P_0)$ and its derivatives can be consistently estimated by the following algorithm:

$$\min_{\gamma_1,\gamma_2} \sum_{i\leq N} \left[Y_i - \gamma_1 - \gamma_2 \left(P_i - P_0\right)\right]^2 G\left(\frac{P_0 - P_i}{a_N}\right)$$

where γ_1 is a consistent estimator of $E(Y|P=P_0)$, and γ_2 is a consistent estimator of $\partial E\left(Y \mid P = P_0\right)/\partial P$. G(.) is a kernel function and a_N is the bandwidth. We use a Gaussian kernel and a bandwidth of 0.2 in the estimation. Obviously, this algorithm is equivalent to applying weighted least square at each observation point, using samples in its “neighborhood”.

We estimate E(lnY|P) and E(X|P) with the above procedure. Then we run the double residual regression of lnY-E(lnY|P) on X-E(X|P). This is a simple OLS regression, except we trimmed off the smallest 2% of the estimated propensity scores with a biweight kernel as suggested by Heckman, Ichimura, Todd, and Smith. The result is consistently estimated coefficients of the linear components of the model, β.

Define the nonlinear component residual as $U=\ln Y- \beta X$. Use local linear regression again to estimate E(U|P) and its first derivative. This first derivative is the marginal treatment effect (MTE). The average treatment effect (ATE) is a simple integration of the MTE with equal weight assigned to each $P(Z)=U_s$. However, treatment on the treated and treatment on the untreated use the following weighting functions:

$$h_{TT}(u_s)=\frac{\left[\int_{u_s}^{1} f(p)\,dp\right]}{E(p)}$$

$$h_{TUT}(u_s)=\frac{\left[\int_{0}^{u_s} f(p)\,dp\right]}{E(1-p)}$$

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.

3. Data

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). We briefly describe our use of the CHIP 95 data here. The data are taken from the urban component of the survey, in which 6,928 households and 21,688 individuals in urban areas of eleven provinces were surveyed for 1995. The survey was funded by the Ford Foundation and a number of other institutes.[6] In the data, annual earnings include regular wages, bonuses, overtime wages, in-kind wages, and other income from the work unit. The hourly wage rate is calculated based on the reported number of working hours. The education measure includes seven degree categories, ranging from below elementary school to college. For more details, see Li (2003).

In China, the definition of labor force is limited to ages 16 or above. As a general rule, in the late 1970s, children entered elementary school at age 7 and remained there for 5 years; junior high school and senior high school each required 2 years. Thus, an individual who was born in 1962 and started school at age 7 would be a senior in upper middle school in 1978 facing the choice of going to college or starting to work. We limit all of our samples to individuals born after 1961 in order to avoid the complicating effects of educational policy during the Cultural Revolution, when many youths were sent to the countryside for “rectification” (or “re-education”), and colleges and even middle schools were either closed or nonfunctioning. The upper birth-year cutoff eliminates observations born too late to have entered college in China's education system (for the probit equations) and too late to have completed college (for the wage equations).

Another sample limitation is based on our need for family background information such as parental education and parental income. Thus, our sample is restricted to working individuals who are living in a household with their parents (for the probit equations) and who have positive earnings in 1995 (for the wage equations). As specified in the model, we only include two education groups: 3 or 4-year college and upper middle school.

4. Empirical Results

Propensity to Acquire a College Education

Table 1 presents estimates of the probit for college attendance in the three sample years, 1988, 1995, and 2002. These probit equations are used to generate a propensity score for each observation, which is the predicted probability of college attendance. The frequency distribution of propensities to attend college provide a reduced-form picture of increasing college attendance in China.

The columns (4), (8), (9), and (13) of table 1 report the mean marginal "propensities" attributable to each independent variable. In our sample years 1988, 1995, and 2002, the effect of parental schooling is highly significant, but it becomes quantitatively smaller over time. The marginal impact of a one-year increase in father's education on the probability of a child attending a 4-year college is 2.1 percentage points in 1988, 3.75 percentage points in 1995, but it drops to only 1.72 percentage points in 2002. The impact of mother's education follows the same time pattern. The impact of parental income on college attendance is also significant in all periods. The marginal impact of 1000 yuan/year in combined parental income increasing the probability of attending college is approximately 1.5 percentage points in 1988, 1 percentage point in 1995 and 2002.

Figure 1 shows the frequency distribution of propensity scores for 1988, 1995, and 2002. The first panel shows the distribution for all observations (S=1 and S=0) in 1988 and 1995; the second panel, while the bottom panel shows separate distributions for college attenders and non-college attenders in 1988 and 1995; the bottom panel shows the separate distributions for attenders and nonattenders in 2002. The rightward shift of the distributions reflects increasing college enrollment and is consistent with the nearly 80% growth of the proportion of the urban population with education of college and above between 1988 and 1995 and more than 100% growth by 1999, as documented in our data by in other studies as well (for example, Zhang and Zhao 2002, table 4). In 1988, the frequency distribution of non-attenders is supported over a range of propensity scores from approximately zero though nearly 0.6; 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 attenders 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.

There are some interesting implications of comparing the distributions and their shifts over time. Table 2 shows that in 1988, 20.8% of the sample were college attenders or college graduates and had a propensity score equal to or greater than 0.324. In 1988 14.4% of the non-attenders had scores higher than this value (yet they didn't go to college), while 60.4% of the attenders had scores less than this value (yet they did go to college). Under the maintained hypothesis that the only unobserved heterogeneity is individual comparative advantage to benefit from college, and that all financial constraints are captured in the probit equation from which we derive the propensity scores, then we may infer that nonattenders with high propensity scores do not choose college because they know they have a comparative advantage as high-school graduates. This hypothetically “scholastically disadvantaged” group made up 11.1% of the entire sample in 1988, 16.7% of the 1996 sample, and 17.5% of the sample in 2002; it amounted to 29.3% of nonattenders in 1995 and 47.5% in 2002. These comparisons, under the hypothesis that only comparative advantage accounts for unobserved heterogeneity is consistent with improved sorting over the years.

College Education and Earnings

Tables 4 and 5 contains the results of OLS, IV, and SPIV estimation of the effect of college attendance on earnings. Table 4 reports the results of benchmark estimates of wage equations in which no variables represented student ability are included as regressors. The benchmark OLS estimates for 1988 and 1995 are commensurate with those reported elsewhere for comparable time periods and increase gradually through 2002(See Fleisher and Wang, 2004, for estimates and a summary of other studies)[7]. The IV estimates of the return to college education (all of which use the propensity score as the instrument for college attendance) are considerably higher than the OLS estimates in the benchmark regressions.

Estimates based on regressions containing a proxy for student ability are reported in Table 5. When either parental education or parental income variables are used to proxy for ability, the OLS estimates of the return to schooling are approximately equal to those reported in Table 4, with the exception of the estimate reported by Heckman and Li (2004), for the year 2000. Their OLS estimated return, with parental income used as an ability proxy, is much higher than their benchmark OLS estimate; however it is about the same as the OLS estimates we obtain from the CHIP data for the years 1988 and 1995 in both the benchmark formulations and in the regressions that include an ability proxy. When parental education is used as a proxy for student ability in the IV earnings equations, the estimated coefficient for college attendance is much higher than the OLS estimates for the years 1988 and 1995, and 2002. However, when parental income is used as a proxy for ability, the IV estimates are approximately the same as the OLS estimates in 1988 and 1995, but higher in 2002 (although much lower than when parental education is the ability proxy)[8].

We turn now to our estimates of returns to schooling based on SPIV estimation. The distinguishing feature of the SPIV procedure is the capacity to retrieve estimates of the marginal treatment effect (MTE) of college education that allow for heterogeneity in the return to schooling. Figures 2 through 3 depict the MTE of college education in China for the years 1988, 1995, 2000, and 2002. Figure 2 compares the MTE from two specifications of the wage equation for 1988 and for 1995. Figure 2a places the 1988 and 1995 MTE curves together so that the effect of including an ability proxy can be seen more clearly. Inclusion of an ability proxy in the local linear regressions simply results in a parallel upward shift of the MTE curve. The shape is not affected across the *Us* dimension.

The heterogeneity model postulates that those who attend college do so because they benefit more than those who choose note to attend. It is important to emphasize that this assumption does imply decisions 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 (CHV 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.* The MTE curves for 1988 support this hypothesis, but they are inconsistent with it in 1995 and, dramatically so, in 2002. The 1995 MTE curves reach a minimum in the middle of the *Us* range and then curve back up toward larger values of *Us*. The 2002 MTE curves are montonically increasing in *Us.* These shapes are inconsistent with the joint hypothesis that that only comparative advantage accounts for unobserved heterogeneity and that the probit estimates of propensity to attend college fully capture financial constraints in 1995 and 2002. They do not suggest improved sorting according to comparative advantage over the years. They are consistent with another hypothesis: that the wage gain to a college education among those least likely to attend college would be higher than among some individuals more likely to attend. They are consistent some barrier to college attendance in China other than lack of ability to benefit financially, e.g. psychic costs or unobserved financial barriers (CHV 2004, p. 25).

4. Conclusion

- The OLS return to college education increased between 1988 and 1995, but increased sharply between 1995 and 2002. In the year 2000, it remained somewhat small by international standards, approximately 7.1% per year of college.
- IV estimates of the return to college are sensitive to the use of a proxy for ability. When parental income is used as a proxy for ability in the local nonlinear regression wage regression, IV estimated returns to college were unchanged between 1988 and 1995 but more than doubled between 1995 and 2002. When parental schooling is used as a proxy for ability, the IV estimated returns to college are higher than when parental income is used and decreased between 1988 and 1995 and increased sharply between 1995 and 2002.
- The time pattern of the average treatment effect (ATE) of college education is similar to that of the IV estimates. In terms of the percentage return per year of college, it is 12.8% in 1988, 11.85 in 1995, and 23.2% in 2002.
- When parental education is used as a proxy for ability, the estimate of heterogeneous return to college for college attenders (TT) falls from 27.6% in 1988 to 13.3% in 1995 and then rises to 16.1% in 2002. The counterfactual return to college attendance for those who did not attend (TUT) rises substantially, from 7.8% in 1988 to 10.6% in 1995, and to 32.3% in 2002.
- When parental income is used as a proxy for ability, TT is smaller in all three years and also declines substantially between 1988 and 2002; similarly, TUT is always smaller when parental income is used as the ability proxy and rises.

- Sorting gain declines substantially, becoming negative in 2002. This evidence is consistent with the increasing importance of unmeasured financial constraints on college attendance and is the crux of our continued research.

.

References

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

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

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

Fan, J. (1992), Design-adaptive Nonparametric Regression, Journal of the American Statistical Associatioin 87: 998-1004.

_____ (1993), Local Linear Regression Smoothers and Their Minimax Efficiencies, The Annals of Staistics 21: 196-216.

Fleisher, Belton M., and Chen, Jian, (1997) "The Coast-Noncoast Income Gap, Productivity, and Regional Economic Policy in China." *J. Comp. Econ* 25, 2: 220-36, 1997.

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

Fleisher, Belton, Sabirianova, Klara, and Wang, Xiaojun (2004), Returns to Skills and the Speed of Reforms: Evidence from Centeral and Eatern Europe, China, and Russia, Ann Arbor, MI: William Davidson Institute Working Paper Number 703.

.Fleisher, Belton M. and Wang, Xiaojun (2001), Efficiency Wages and Work Incentives in Urban and Rural China, Journal of Comparative Economics 29(4): 645-662.

________________________(2004a), Skill Differentials, Return to Schooling, and Market Segmentation in a Transition Economy: The Case of Mainland

China (with Xiaojun Wang) *Journal of Development Economics* 73(1), 715-728.

Fleisher, Belton and Wang, Xiaojun (2004b), Higher Education in China: A Growth Paradox First Global Conference in Honor of Gregory Chow. Aldershot, UK: Ashgate Publishing Ltd. (Forthcoming).

Fleisher, Belton and Wang, Xiaojun (2004c), Returns to Schooling in China under Planning and Reform," William Davidson Institute Working Paper No. 704, http://www.bus.umich.edu/KresgeLibrary/Collections/Workingpapers/wdi/wp704.pdf

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

Giles, John, Park, Albert, and Zhang, Juwei (2004), The Great Proletarian Cultural Revolution, Disruption to Education, and Returns to Schooling, East Lansing, Michigan: Department of Economics, Michigan State University.

Hannum, Emily and Wang, Meiyan (2004), Geography and Educational Inequality in China, working paper, Department of Sociology and Population Studies Center, University of Pennsylvania.

Heckman, James J. (2004), China's Human Capital Investment, China Economic Review (in press).

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

Heckman, James J. and Li, Xuesong (2003), Selection Bias, Comparative Advantage, and Heterogeneous Returns to Education: Evidence from China in 2000.

Jones, Derek C. and K. Ilayperuma (1994), Wage Determination under Plan and Early Transition: Evidence from Bulgaria, Working Paper No. 94/7, Hamilton, NY: Department of Economics, Hamilton College.

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

Li, Haizheng (2003), Economic Transition and Returns to Education in China, Economics of Education Review 22 (317-328).

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

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

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

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

Munich, Daniel, Svejnar, Jan, and Terrell, Katherine (2000), Returns to Human Capital under the Communist Wage Grid and During the Transition to a Market Economy, paper presented at the World Conference of the European Association of Labor Economists and Society of Labor Economists, Milan, Italy.

National Bureau of Statistics of China (1994), Annual Population Change Survey 1993, Beijing, China.

National Bureau of Statistics of China (2002), 2000 Population Census, Beijing, China.

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

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

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

Yang, Dennis Tao (2004), Determinants of Schooling Returns during Transiton: Evidence from Chinese Cities, Blacksburg VA: Department of Economics, Virginia Polytechnic Institute and State University.

Zhang, Junsen and Zhao, Yaohui (2002), Economic Returns to Schooling in Urban China, 1988-1999, paper presented at the 2002 meetings of the Allied Social Sciences Association, Washington DC. Hong Kong; Chinese University of Hong Kong, jszhang@cuhk.edu.hk.

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

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

Zhou, Xueguang and Moen, Phyllis (2002), 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]

Table 1 Propensity Estimates

	CHIP88				CHIP95				H&Li (2000)	Chip02			
Variable	**Parameter**	**t-ratios**	**p-values**	**(4) Mean Marginal Effect**	**Parameter**	**t-ratios**	**p-values**	**(8) Mean Marginal Effect**	**(9) Mean Marginal Effect**	**Paramter**	**t-ratio**	**p-values**	**(13) Mean Marginal Effect**
CONST	-1.3255	-5.7944	0	**-0.3514**	-1.9984	-7.9504	0	**-0.7868**		-0.985	-4.5055	0	**-0.3677**
FEDU	0.0802	5.2656	0	**0.0213**	0.0953	6.2347	0	**0.0375**	**0.0211**	0.0462	3.6987	0.0001	**0.0172**
MEDU	0.0159	1.0198	0.154	**0.0042**	0.0718	4.8196	0	**0.0283**	**0.0126**	0.0448	3.4419	0.0003	**0.0167**
FWAGE	0.0302	0.9124	0.1809	**0.008**	0.0161	0.6243	0.2663	**0.0063**	**0.0040***	0.0132	2.3574	0.0093	**0.0049**
MWAGE	0.0889	1.9831	0.0238	**0.0236**	0.0418	1.2254	0.1104	**0.0165**		0.0139	1.897	0.029	**0.0052**
CHL	-0.1287	-2.4677	0.0069	**-0.0341**	-0.043	-0.6533	0.2569	**-0.0169**		-0.0471	-0.5598	0.2878	**-0.0176**
SEX	-0.0082	-0.0967	0.4615	**-0.0022**	-0.1196	-1.3953	0.0816	**-0.0471**		-0.3658	-5.2193	0	**-0.1366**
ETHNIC	-0.0247	-0.1054	0.458	**-0.0066**	-0.2063	-0.9342	0.1752	**-0.0812**		-0.1652	-1.1316	0.129	**-0.0617**
BY1968	-0.9831	-6.4701	0	**-0.2606**	-0.2089	-1.4676	0.0713	**-0.0823**		0.4368	2.5557	0.0053	**0.1631**
BY1967	-0.5369	-3.2247	0.0006	**-0.1423**	0.1256	0.7025	0.2413	**0.0495**		0.3796	1.4943	0.0676	**0.1417**
BY1966	-0.4494	-2.756	0.003	**-0.1191**	0.0507	0.2882	0.3866	**0.02**		0.4366	1.5641	0.059	**0.163**
BY1965	-0.4114	-2.536	0.0057	**-0.1091**	0.032	0.1823	0.4277	**0.0126**		0.2397	0.8472	0.1985	**0.0895**
BY1964	-0.2138	-1.4434	0.0746	**-0.0567**	0.0489	0.2662	0.3951	**0.0192**		0.4112	1.3665	0.086	**0.1535**
BY1963	-0.2327	-1.503	0.0665	**-0.0617**	0.0479	0.2253	0.4109	**0.0188**		-0.1613	-0.4929	0.3111	**-0.0602**
BY1962					0.3006	1.3692	0.0856	**0.1182**		0.2677	0.8435	0.1996	**0.0999**
	FLWdiary4.txt BW 0.2				Flw95output BW 0.2								FLW02out1212.txt

Notes: The dependent variables is a dummy variable = 1 for graduated from college. The independent variables are, respectively, father's education in years, mother's education in years, mother's and father's annual income in 1000 *yuan* per year, including cash and in-kind benefits, number of children in family of origin, a dummy variable = 1 if respondent is male, dummy variable =1 if ethnicity is not Han Chinese, and dummy variables for birth year.

*The coefficient is for the variable parental income.

Table 2 Comparison of Propensity Distributions

	1988	1995	2002
Proportion of sample who are college attenders or graduates	21%	42.7%	63.0%
Cutoff Propensity	.324	.480	.588
Proportion of attenders with scores less than the cutoff	60.4%	39.9%	27.6%
Proportion of attenders with low scores as proportion of sample	12.6%	17.0%	17.4%
Proportion of nonattenders with scores greater than the cutoff	14.4%	29.3%	47.5%
Proportion of nonattenders with high scores as proportion of sample	11.1%	16.7%	17.5%

Note: The cutoff percentage is the propensity score that corresponds to the cumulative frequency of the total sample that were attending or had graduated from college in the sample year.

Table 4 Benchmark regression estimates and treatment effect estimates

Parameter	CHIP88	CHIP95	CHIP02	H&Li (2000)
OLS	.1986	.2307	.3142	.0856
IV	.3435	.3724	.9812	.2192
ATE	.2556	.3473	.8248	.2321
TT	.7868	.3883	.3943	.1909
TUT	.1147	.3135	1.4958	.2679
Bias = OLS-ATE	-0.0569	-.1166	-.5106	-.1465
Selection Bias = OLS-TT	-0.5882	-.1576	-.0800	-.1053
Sorting Gain = TT-ATE	0.5312	0.0410	-.4306	-0.0412
TT-TUT	0.672	0.0748	-1.102	-0.077
Source	FLWdiary4.txt BW 0.2	Flw95output BW 0.2	Flw02out 1212	Table 7a w/o parental inc

Note: Dependent variable is monthly wage in 1988, hourly wage in 1995 and 2002. OLS regressors are a dummy 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 predicted college attendance based on the propensity score as an instrument. The treatment effect estimates are based on results from local linear regression.

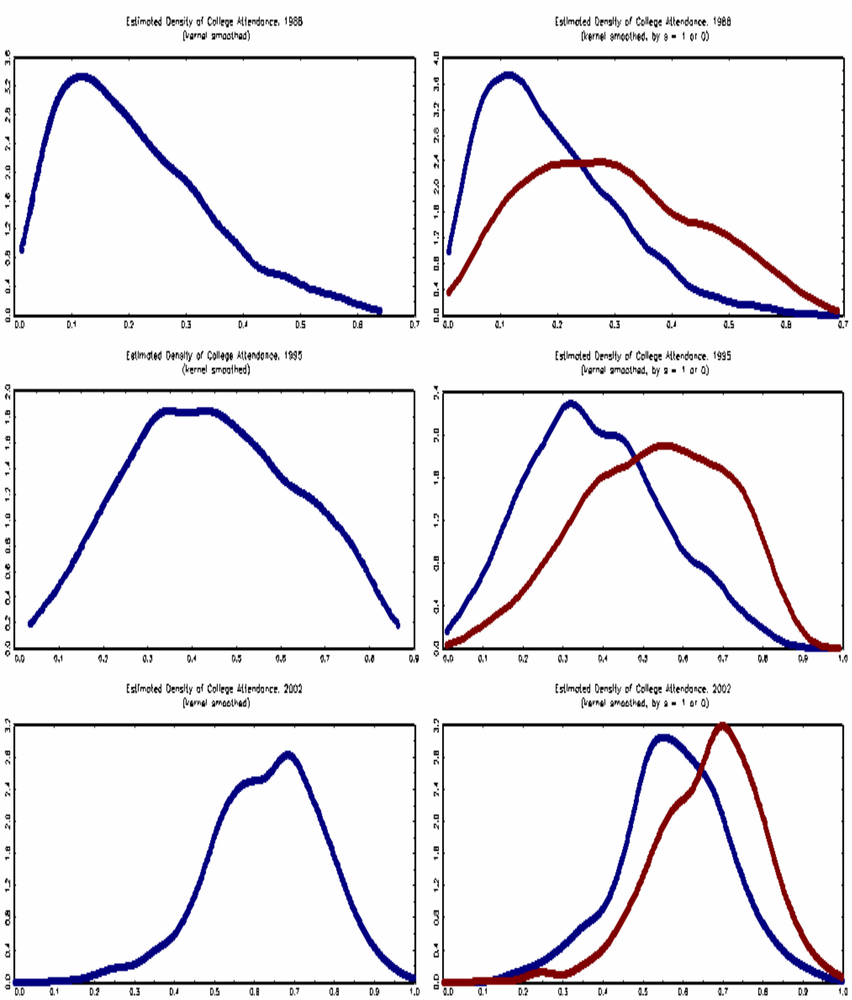
Table 5 Regression estimates with ability proxy included and treatment effect estimates

Parameter	CHIP88	CHIP88	CHIP95	CHIP95	CHIP02	CHIP02	H&Li (2000)
Ability proxy	fedu medu	fwage mwage	fedu medu	fwage mwage	fedu medu	fwage mwage	parental income
OLS	0.2029	0.1985	.2127	.2114	.2814	.2687	.2929
IV*	0.8494	0.2033	.5963	.1995	1.4711	.4764	.5609
ATE	0.6239	0.1854	.5660	.1889	1.3044	.4084	.4336
TT	1.6530	0.5817	.6460	.2215	.8168	.2025	.5149
TUT	0.3510	0.0804	.5002	.1621	2.064	.7293	.3630
Bias = OLS-ATE	-0.4211	0.0130	-.3533	.0226	-1.023	-.1397	-.1407
Selection Bias = OLS-TT	-1.4502	-0.3832	-.4333	-.0100	-.5354	.0662	-.2220
Sorting Gain = TT-ATE	1.0291	0.3963	.0800	.0326	-0.4876	-.2059	.0813
TT-TUT	1.302	0.5013	0.146	0.0594	-1.25	-.527	0.155
Source	FLWdiary4.txt BW 0.2	FLWdiary4.txt BW 0.2	Flw95output BW 0.2	Flw95output BW 0.2	Flw02out 1212	Flw02out 1212	Table 6 w(?) parental inc BW 0.3 or 0.4??

Note: Dependent variable is monthly wage in 1988, hourly wage in 1995 and 2002. OLS regressors are a dummy 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 predicted college attendance based on the propensity score as an instrument. The treatment effect estimates are based on results from local linear regression.

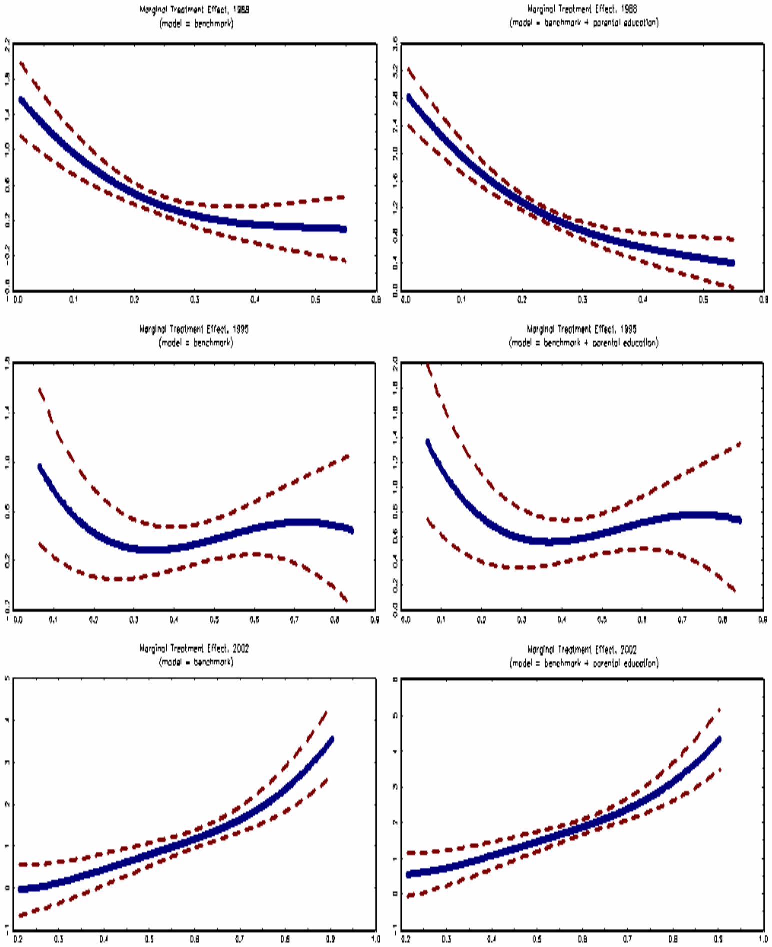
From Figure 1.11_30 and Prob02a.doc

Figure 1 Propensity to Attend College Frequency Distributions 1988 and 1995



From Figure 2b.doc

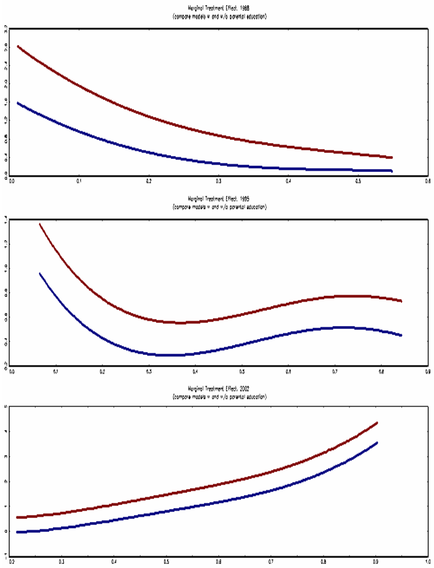
Figure 2 Marginal Treatment Effects 1988, 1995, and 2002



Left side are benchmarks, which do not contain ability proxies. Right side include parental education as ability proxy.

Figure 3.doc

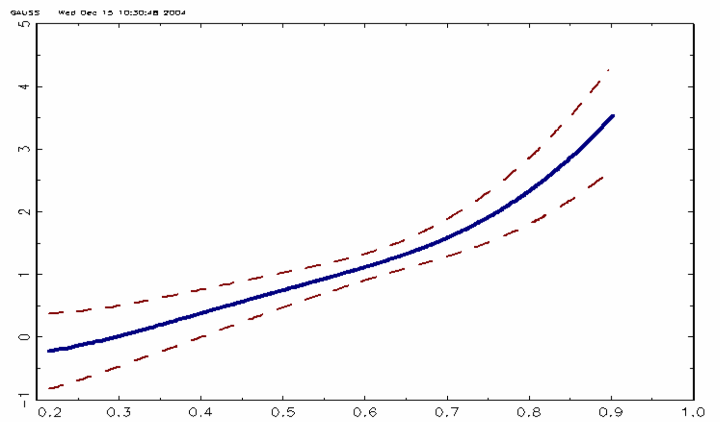
Figure 3 MTE Curves with and without ability proxies (parental schooling)



Upper curves reflect inclusion of parental schooling as ability proxies.

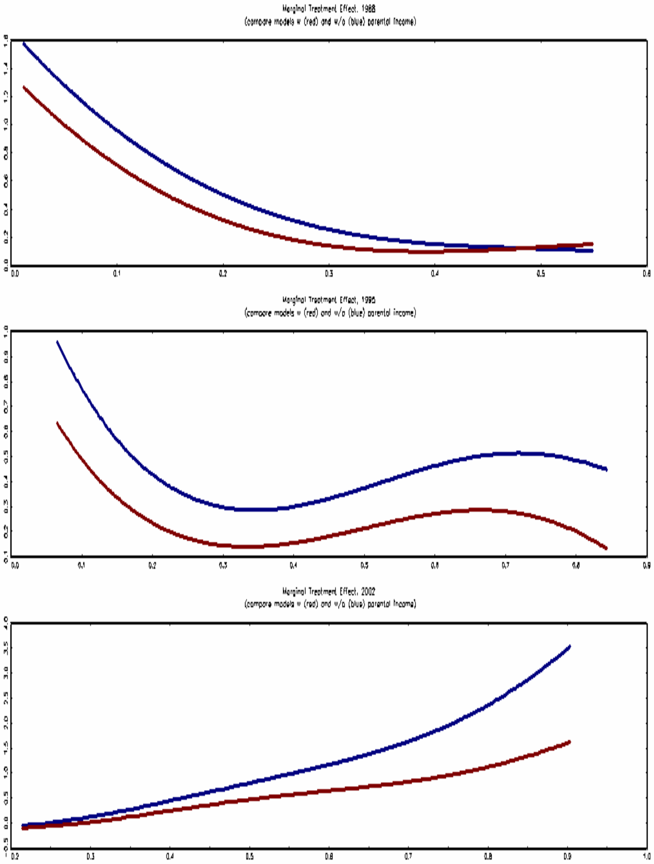
From mte_hscr.doc

Figure 4 MTE 2002 Ability Proxy Rank in High School



From figure 3b.doc

Figure 5 MTE 2002 with Parental Income as Ability Proxy



Benchmark curves are on top

Methodology Appendix

The marginal treatment effect (MTE) and its derivative statistics derived from it are estimated using the method developed in Heckman, Ichimura, Todd, and Smith (1998) and Fan (1992 and 1993). These derivative statistics include the average treatment effect (ATE), treatment of the treated (TT), treatment of the untreated (TUT), bias, selection bias, and sorting gain.

We specify the following model of income and schooling choice:

$$\ln Y_1 = \mu_1(X, U_1)$$
$$\ln Y_0 = \mu_0(X, U_0)$$

where a subscript indicates whether the individual is in the schooled state (1) or the unschooled state (0). Y is income, X is observed heterogeneity, and U is unobserved heterogeneity in wage determination. In general, the functional forms can have a nonlinear component, and $U_1 \neq U_0$.

The schooling choice comes from the following latent dependent model:

$$S^* = \mu_s(Z) - U_s$$
$$S = 1 \ \ if \ \ S^* \geq 0$$

where S^* is a latent variable whose value is determined by an observable component $\mu_s(Z)$ and an unobservable component U_s.

In our empirical work, Z is a vector of variables that predict the probability of attending college. It includes parental education, parental income, number of children, gender, ethnic group, and birth year dummies. X is a vector that holds explanatory power on wages. In the benchmark setting, this includes experience, experience squared, gender, ethnic group, ownership, industry, and location.

In the first step, a probit model is used to estimate the $\mu_s(Z)$ function. The predicted value is called the propensity score, $\hat{P}_i$, where the subscript i indicates each individual.

The second step adopts a semi-parametric procedure in which *local linear regressions* are used frequently. Fan (1992, 1993)[9] develops the distribution theory for the local linear estimator of $E(Y|P=P_0)$, where Y and P are random variables. $E(Y|P=P_0)$ and its derivatives can be consistently estimated by the following algorithm:

$$\min_{\gamma_1,\gamma_2}\sum_{i\leq N}\left[Y_i-\gamma_1-\gamma_2\left(P_i-P_0\right)\right]^2 G\left(\frac{P_0-P_i}{a_N}\right)$$

where γ_1 is a consistent estimator of $E(Y|P=P_0)$, and γ_2 is a consistent estimator of $\partial E\left(Y\mid P=P_0\right)/\partial P$. G(.) is a kernel function and a_N is the bandwidth. We use a Gaussian kernel and a bandwidth of 0.2 in the estimation. Obviously, this algorithm is equivalent to applying weighted least squares at each observation point, using samples in its “neighborhood”.

We estimate E(lnY|P) and E(X|P) with the above procedure. Then we run the double residual regression of lnY-E(lnY|P) on X-E(X|P). This is a simple OLS regression, except that we trim off the smallest 2% of the estimated propensity scores with a biweight kernel as suggested by Heckman, Ichimura, Todd, and Smith. The result is consistently estimated coefficients of the linear components of the model, β.

Define the nonlinear component residual as U=lnY- βX. Use local linear regression again to estimate E(U|P) and its first derivative. This first derivative is the marginal treatment effect.

The average treatment effect is a simple integration of the MTE with equal weight assigned to each $P(Z)=U_s$. However, treatment on the treated and treatment on the untreated use the following weighting functions:

$$h_{TT}(u_s) = \frac{\left[\int_{u_s}^{1} f(p)\,dp\right]}{E(p)}$$

$$h_{TUT}(u_s) = \frac{\left[\int_{0}^{u_s} f(p)\,dp\right]}{E(1-p)}$$

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.

---

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

[*] Corresponding author.

[2] For China See Zhang and Zhao (2002), Li (2003), and references cited in Fleisher and Wang (2001). For other countries, see Munich, Svejnar, and Terrell (2000), Orazem and Vodopivec, 1995, and Jones and Ilayperuma, 1994.

[3] *Econometrica* 66, 5 (Sept. 1998): 1017-1098.

[4] These derivatives include average treatment effect (ATE), treatment on the treated (TT), treatment on the untreated (TUT), bias, selection bias, and sorting gain.

[5] Fan (1992): *Journal of the American Statistical Association* 87: 998-1004. Fan (1993): *The Annals of Statistics* 21: 196-216.

[6] The CHIP-95 data are available to the public at the Inter-university Consortium for Political and Social Research (ICPSR).

[7] The OLS estimate of return to schooling in 2000 reported by Heckman and Li (2004) is problematic. In their benchmark regression, they report an OLS estimate of .0856 for four years of college, implying an annual rate of return of

only 2.1%, which is much lower than estimated returns in 1988 and 1995 ; in an OLS regression that includes parental income as a proxy for ability, they report an estimate of .2929 for four years of college, implying an annual rate of return of 6.6%, about the same as in 1988 and 1995. Moreover, the OLS estimates reported by Heckman and Li (2004) are low in comparison to those obtained in other research. Giles, Park, and Zhang (2004) use data for the year 2000 obtained from the China Urban Labor Survey conducted in 2001. The data cover the cities of Fuzhou, Shanghai, Shenyang, Wuhan, and Xian. Using these data, they obtain an estimate for return to four years of college education of approximately 0.52, which converts to approximately11% annual rate (personal conversation with John Giles).

[8] Heckman and Li (2004), however, report an IV estimate of return to schooling equal to 0.5609 for college graduates based on a regression in which parental income is used as a proxy for ability. This is nearly twice as large as their reported OLS estimate.

[9] Fan (1992): *Journal of the American Statistical Association* 87: 998-1004. Fan (1993): *The Annals of Statistics* 21: 196-216.