Human capital and wages: controlling for measurement errors, unobserved heterogeneity and endogeneity. Application to China*
Preliminary and incomplete

Jean-Louis Arcand† Kelly Labar‡

March 22, 2007

Abstract

We consider the effect of human capital (health and education), on the wages of a panel of workers from the 1991, 1993, 1997, 2000 and 2004 China Health and Nutrition Surveys. We implement an estimator that simultaneously corrects for measurement error using the "internal" instruments of Dagenais and Dagenais, 1997, correlated individual effects, selection bias and endogeneity (Wooldridge and Semikyna, 1995). Preliminary results show that neither health nor education have an impact on wages.

Keywords: health, wages, education, China

JEL: J31, I21, C30

1 Introduction

Since the end of the 1970s, China has embarked on a momentous process of economic and social transition. This has had major consequences on the functioning of the labor market, which has become increasingly flexible and has seen growing private sector employment. The process underway involves changes both in recruitment methods and in the determination of wages. The latter should increasingly translate worker productivity and reward different forms of human capital.

1.1 The determinants of wages in China

Before the reform process, wages were essentially determined by an overarching concern for equity. Set by the government, salaries were low but homogeneous, with change being largely a function of seniority. Wage earners also received numerous in-kind benefits, such as subsidized housing

*Corresponding author: Kelly Labar, CERDI-CNRS, Université d'Auvergne, 65 boulevard François Mitterrand, 63000 Clermont Ferrand, France; tel: +33-(0)4 73177400, fax: +33-(0)4 73177428.

†CERDI-CNRS, Université d'Auvergne, and European Development Network.

‡CERDI-CNRS, Université d'Auvergne.

and medical care. Worker mobility was very limited, leading to important disequilibria in various segments of the labor market (Chen, Demurger, and Fournier (2005), Li and Zax (2004)).

Conventional wisdom has it that the reforms have radically modified the structure of Chinese labor markets. The managers of state-owned enterprises have seen their managerial autonomy increase. Cash bonuses have appeared, and often reward employees perceived as being particularly productive (although the structure of many bonus-schemes is still largely driven by concerns of equity). Concomitantly, a private sector has emerged, which accounted for 35 % of total employment in 1995. Heterogeneity has also appeared in terms of ownership structures, leading to significant differences in wages. For example, Chen, Demurger, and Fournier (2005), who identify six distinct ownership structures (state-owned enterprises (SOEs), local publicly owned enterprises, urban collective enterprises, private or individual enterprises and foreign enterprises), have highlighted the heterogeneity in employee characteristics and wages, based on the ownership structure of the employer. Employees in SOEs, for example, tend to be older, more experienced and better qualified than their brethren working for other types of enterprises. Foreign enterprises, in contrast, provide the highest level of wages, but demand significantly lengthier workhours.

As ownership structures have evolved, gender differences have emerged, and significant cleavages have also appeared between recent migrants from the countryside and native city dwellers. Maurer-Fazio, Rawski, and Zhang (1999) show that, despite constitutional guarantees, the fundamentally patriarchal nature of Chinese society has led to women being increasingly shunted into low wage occupations, the result being a 50% gender wage gap, which also varies according to ownership structure. Interestingly, the gender wage gap is not systematically lower in publicly-owned enterprises. Collectives do however, continue to have the most homogeneous wage structure. Migrants, for their part, suffer from a 50% wage gap with respect to native city-dwellers and work 14 more hours per week, as shown by Meng and Zhang (2001).

As a result of the reforms and the increasing role played by the private sector, greater competition should lead to various forms of human capital being more faithfully rewarded in terms of wages. This intuition is confirmed by Liu (2001), who compared the returns to education in a market-oriented province (Guangdong) with those in a province (Liaoning) where central planning still plays a major role. The societal downside is that inequality in education will tend to be increasingly translated into inequality in terms of wages.

Consequently, the role of human capital, which can be divided between health and education, in the determination of wages is expected to increase with the movement of the economy toward a relatively market.

1.2 Human capital and wages

The link between human capital and wages can be measured by the returns to education and health.

The consequences of the reforms on the returns to education are considered by Li (2003), who studies differences in the returns to education between cohorts who started working during different periods. For those who began working before 1979, the returns to an additional year of education was equal to 4.7%, while the corresponding figure for those who began working between 1980 and 1987 was 7.3%, the difference being statistically significant; for those who began working between

1988 and 1995, the figure is 6,5% (the difference between these two last values was not found to be statistically significant). The flavor of these results is confirmed by Maurer-Fazio, Rawski, and Zhang (1999), who find that the returns to education are higher in the private sector. Maurer-Fazio and Dinh (2002) show that disparities between men and women, as well as between migrants and native city-dwellers are also important in terms of the returns to education.

While the educational component of human capital would appear, in light of the available empirical evidence, to play an increasingly important role in determining wages in China, little is known concerning other forms of human capital, the most important among these being health.

The mechanisms whereby health affects the productive capacity of a worker and consequently (if individuals are paid according to their marginal productivity) wages are not difficult to fathom. Nevertheless, in order to study the importance of the phenomenon empirically, it is necessary to measure health as accurately as possible. In the empirical litterature dealing with the link between health and productivity, four measures are typically used: height, body mass index (BMI, weight expressed in kilograms divided by the square of height expressed in metres), as well as calorie and protein intake.

In this paper we shall, for the time being, restrict our attention to the first measure. Nobel Laureate Robert Fogel (2004) refers to this indicator as an element of "physiological capital". Height, thought to be particularly sensitive to early childhood nutritional and health status, conveys information that may proxy more general living standards (Schultz (2001), Strauss and Thomas (1998), Steckel (1995), Fogel (2004)). It is also used by some authors as an indicator of the physiological modifications that arise as a result of changes in individual behaviour over time (Steckel (1995), Fogel (1994)). It is also a good proxy for physical strengh, which may be important in certain sectors.

In the development economics literature, the relationship between height and wages has been shown to be positive and statistically significant in Brazil (Thomas and Strauss (1997)), Ghana and Côte d'Ivoire (Schultz (2001)), as well as in the Philippines (Haddad and Bouis (1991)), the causal link running from height to physical strength, and thence from perceived physical strength to wages, at least in occupations requiring significant inputs of physical effort.

1.3 Empirical issues

The estimation of wage equations that include education and height is plagued by the classical problem of individual heterogeneity. While educational attainment may reflect unobserved individual ability, height is in large part determined by an individual's genetic heritage. In order to interpret the effect of these variables on wages as being that of contemporaneous health, estimates need to be purged of their genetic component. In a number of studies, this is done by relying on a sample of monozygotic twins (Behrman and Rosenzweig (2001)).

Here, we exploit the panel nature of five rounds of the China Health and Nutrition Survey to the same effect. Dealing with individual-specific unobserved heterogeneity has been surprizingly rare in studies of the returns to health in developing countries, rendering their results suspect (Thomas and Strauss (1997) and Schultz and Tansel (1997) are cases in point). When individual effects are purged, the consequences with respect to estimates based on pooling are mixed. In the Indian example considered by Deolalikar (1988), controlling for unobserved heterogeneity confirms and strengthens the pooling results. In the Philippino case considered by Haddad and Bouis (1991), the effect of health on wages vanishes once individual effects are controlled for.

A second problem involves the potential for significant selection bias in the estimation of the wage equations. If the probability of wage employment is determined in part by attributes such as height or education, one runs straight into the type of selection problems first considered by Heckman. Moreover, "within" estimation of the wage equation using individuals who are employed over several periods will usually not solve this problem, as there is no reason for the underlying selection bias to be time-invariant. As such, estimates of the impact of education and height on wages will be biased. In order to solve this problem, we consider exclusion restrictions based on ethnically-specific "one child" policies, as well as newspaper availability.

The third potential problem involves the usual concern with omitted variables. Intrinsic ability, which is unobserved, is subsumed in the residuals. To the extent that individual fixed effects do not account for all of this portion if the disturbance term, our estimates will still be biased.

In particular, while height is likely to be predetermined, it may be measured with error, and this may bias key parameter estimates if measurement error is not random (aside from the usual attenuation bias stemming from a classical errors in variables problem). In this paper, we make use of the internal instrument set suggested by Dagenais and Dagenais (1997), which is geared explicitly towards dealing with measurement error, in the absence of additional exogenous instruments.

2 Selection bias, unobserved heterogeneity and endogeneity

The statistical model considered here is given by the following equation :

$$\ln w_{it} = x_{it}\beta + \alpha_i + \varepsilon_{it}, \tag{1}$$

Following Mundlak (1978), the selection equation is given by :

$$s_{it} = \mathbf{1}\left[z_{it}\gamma_t + z_{i\bullet}\psi_t + \upsilon_{it} > 0\right], t = 1, 2, 3, 4, 5 \tag{2}$$

where $i = 1, ..., N$ represent individuals, $t = 1, 2$ represent periodes, $\mathbf{1}[.]$ is an indicator function that is equal to 1 when the inequality in square brackets is satisfied and individual i is a wage-earner at time t and is equal to zero otherwise, w_{it} represents the hourly wage rate, x_{it} is our matrix of explanatory variables in the wage equation, z_i is the matrix of exogenous variables containing instruments and exogenous variables included in x_{it}. These last variables will be used for the identification in the selection equation and for the estimation of the wages equation. γ_t is the coefficient vector associated with the variable z_{it},which can differ by period, $z_{i\bullet} = \frac{1}{2}(z_{i1} + z_{i2})$ is the individual specific mean of the explanatory variables contained in the selection equation. α_i are the unobserved individual effects in the wage equation and ε_{it} and υ_{it} are disturbance terms assumed to be jointly distributed according to the normal density.

We assume that there are exclusion restrictions for the wage equation (x_{it} is a sub-sample of z_{it}) so as not to base identification solely on the non linearity of the inverse Mills ratio. We consider individuals observed over at least two periods wether if they are wage earners or not.

2.1 Testing for non random selection

To test for the presence of non random-selection, Semykina and Wooldridge (2005) propose in a first step to estimate the probit for equation (2) on the sub-periods independently. After recovering the Inverse Mills Ratio (IMR) $\widehat{\lambda}_{it}$, we esimate (using a FE-2SLS procedure) the following equation:

$$\ln w_{it} = x_{it}\beta + \rho_1 \widehat{\lambda}_{it} + \varepsilon_{it}, \tag{3}$$

where the instruments are given by z_{it} and $\widehat{\lambda}_{it}$.

2.2 Correction for non-random selection, unobserved heterogeneity and endogeneity in panel data

To estimate β in equation (1) in a consistent way, correcting simultaneously for selection bias, unobserved individual effects and endogeneity, we use the same IMR as above, and estimate the following equation

$$\ln w_{it} = x_{it}\beta + z_{i\bullet}\delta + \rho_2 \widehat{\lambda}_{it} + \varepsilon_{it}, \tag{4}$$

by 2SLS over the pooled data, where the instruments are given by $z_{it}, z_{i\bullet}$ et $\widehat{\lambda}_{it}$. Finally, given that our specification includes a generated regressor ($\widehat{\lambda}_{it}$), we bootstrap all standard errors.

3 The higher moments instrument set of Dagenais and Dagenais (1997)

The estimator used here in order to deal with the measurement errors problem is inspired by Dagenais and Dagenais (1997), where the matrix of feasible instruments, denoted by $Z_{it} = (z_{1it}, z_{2it}, z_{3it}, z_{4it}, z_{5it}, z_{6it}, z_{7it})$, is given by:

$$\begin{aligned}
z_{1it} &= x_{it} * x_{it}, \\
z_{2it} &= x_{it} * y_{it}, \\
z_{3it} &= y_{it} * y_{it}, \\
z_{4it} &= x_{it} * x_{it} * x_{it} - 3x_{it}\left(\frac{x'_{it}x_{it}}{N} * I_r\right), \\
z_{5it} &= x_{it} * x_{it} * y_{it} - 2x_{it}\left(\frac{x'_{it}y_{it}}{N} * I_r\right) - y_{it}\left[\iota'_r\left(\frac{x'_{it}y_{it}}{N} * I_r\right)\right], \\
z_{6it} &= x_{it} * y_{it} * y_{it} - x_{it}\left(\frac{y'_{it}y_{it}}{N}\right) - y_{it}\left(\frac{y_{it}x_{it}}{N}\right), \\
z_{7it} &= y_{it} * y_{it} * y_{it} - 3y_{it}\left(\frac{y'_{it}y_{it}}{N}\right),
\end{aligned}$$

where y_{it} is the dependent variable, and where the symbol $*$ designates the Hadamard element-by-element matrix multiplication operator, I_r is an r-dimensional identity matrix, and ι_r is an $r \times 1$ vector of ones. Detailed proofs of the orthogonality of these instruments with respect to the disturbance term are provided in Dagenais and Dagenais (1997).

The resulting "higher moments" estimator, which we shall denote by β_H, is consistent when there are EV and is much less erratic than other estimators based on sample moments of order higher than two heretofore suggested in the literature. Note that various implementations of the proposed instrument set are possible. These include Fuller, GMM (the road taken in an earlier paper by Dagenais and Dagenais (1994)), Nagar (or bias-adjusted 2SLS, see Donald and Newey (2001)), or general k-class estimation.

4 Data and descriptive results

4.1 Data

The data used in this paper stem from the 1991, 1993, 1997, 2000 and 2004 rounds of the China Health and Nutrition Survey (CHNS). The CHNS is an ongoing longitudinal survey that covers eight provinces (Guangxi, Guizhou, Henan, Hubei, Hunan, Jiangsu, Liaoning, Heilongjiang and Shangdong). Although the survey is not nationally representative, these provinces were selected to provide significant variability in geography, economic development and health indicators, so that they may be considered to be generally representative of all provinces in the country.

A multistage random-cluster sampling procedure was used to draw the sample from each of the provinces. Counties in the eight provinces were stratified by income (low-, middle- and high-income tertiles) with per capita income figures from the State Statistical Office, and a weighted sampling scheme was used to randomly select four counties in each province (one low income, two middle income, and one high income). Probability-proportional-to-size sampling was used to select the sample from these units. In addition, urban areas initially not within the county-strata were later incorporated by including the provincial capital and a low-income city from each province. Within each county, the township capital was selected and three villages were chosen randomly. Within each city, urban and sub-urban neighbourhoods were randomly selected. The same random selection procedure was used to choose the neighbourhoods for townships and villages.

Anthropometric information was obtained concerning all survey participants, and we only make use of a small subset of the available measurements in this paper. Measurements were carried out by trained health workers who followed standard protocols and techniques. Height was measured without shoes to the nearest tenth of a centimeter with a portable stadiometer. Each of these measurements was done by at least two health workers; one worker took the measurements, and another recorded the reading.

Our basic sample consists of males, of greater than 16 years of age, whose anthropometric measurements fall within the bounds of what is considered reasonable by nutritionists in the Chinese context.

Our control variables include age, the sector in which the individual works, where we differentiate between the public sector (enterprises owned by all public institutions) and the state sector (enterprises owned by the central government only).

Identification of the first-stage probit equation is achieved through two exclusion restrictions. In order to be valid, the variables involved must affect the likelihood of the individual being a wage earner, without having any direct effect on the individual's wage. We consider two variables: the availability of newspapers for the location where individual lives, and the ethnic differentiation and geographic variability in the application of the "one-child" policy. This last variable is defined thanks to two characteristics. The first is the local policy concerning the one child policy, differentiated by ethnic group[1]. In some locations, non-Han Chinese are allowed to have more than one child. That is done in an effort to re-equilibrate the ethnic structure of China. The second characteristic

[1] This variable corresponds to the question: "Are minority couples allowed to have two children no matter what the circumstances?".

is the ethnicity of individuals. It follows that we can construct a variable that is equal to 1 if the individual belongs to an ethnic minority and lives in an area where more than one child is allowed for members of the minority, and 0 otherwise.

These two variables "news-paper availability" and "one child policy" are good identification variables for our selection equation in that the former leads to variations in the cost of job search , while the latter provides exogenous variation in the number of children in the household which, in other contexts, has been proved to significantly affect the likelihood of being a wage earners.

4.2 Descriptive results

Descriptive statistics are given in the table 1 for the full sample and the sample of wage earners.

On average, wage earners are younger and are more likely to work in the public and state sectors than are non-wage earners (the latter may be employed but their remuneration does not take the form of wages). Their educational level is also higher. Wage earners are more likely to live in urban areas.

Considering the within individuals' standard deviation for height and education, we see that they are relatively high considering that the individuals included in our sample have finished their studies and are adults. The substantial variations of height and education over time for a given individual are suggestive of some important measurement error. This is our main reason for applying the Dagenais and Dagenais (1997) estimator.

5 Results

Tables 2 and 3 present the preliminary results of the estimations.

In table 2, we add a column in which we distinguish the effects of height and education before and after the year 1997. We suspect a non linear impact due to the deepening of the reforms in the second part of the period considered.

Those first results show no impact of our interest variables on wages.

References

Behrman, J. R., and M. R. Rosenzweig (2001): "The Returns to Increasing Body Weight," Penn Institute for Economic Research Working Paper 01-052.

Chen, Y., S. Demurger, and M. Fournier (2005): "Wage Differentials and Ownership Structure in Chinese Enterprises," mimeo, CERDI-CNRS, Université d'Auvergne.

Dagenais, M. G., and D. L. Dagenais (1994): "GMM Estimators for Linear Regression Models with Errors in Variables," Cahier 9404, Département de sciences économiques, Université de Montréal.

——— (1997): "Higher Moment Estimators for Linear Regression Models with Errors in the Variables," *Journal of Econometrics*, 76(1-2), 193–221.

Deolalikar, A. (1988): "Nutrition and Labor Productivity in Agriculture," *Review of Economics and Statistics*, 70(3), 406–413.

Donald, S. A., and W. K. Newey (2001): "Choosing the Number of Instruments," *Econometrica*, 69(5), 1161–1191.

Fogel, R. W. (1994): "Economic Growth, Population Theory and Physiology," *American Economic Review*, 84(3), 369–395.

——— (2004): "Health, Nutrition and Economic Growth," *Economic Development and Cultural Change*, 52(3), 643–658.

Haddad, L. J., and H. E. Bouis (1991): "The Impact of Nutritional Status on Agricultural Productivity: Wage Evidence from the Philippines," *Oxford Bulletin of Economics and Statistics*, 53(1), 45–68.

Li, H. (2003): "Economic Transition and Returns to Education in China," *Economics of Education Review*, 22(3), 317–328.

Li, H., and J. Zax (2004): "Economic Transition and Labor Market in China," Working paper.

Liu, Z. (2001): "The Effects of Economic Reforms on Wage Inequality: Some Evidence from China," *Applied Economics Letters*, 8(4), 285–288.

Maurer-Fazio, M., and N. Dinh (2002): "Differential Rewards to, and Contributions of, Education in Urban China's Segmented Labor Market," processed.

Maurer-Fazio, M., T. Rawski, and W. Zhang (1999): "Inequality in the Reward for Holding Up Half the Sky: Gender Wage Gaps in China's Urban Labor Market, 1988-94," *The China Journal*, 41, 55–88.

Meng, X., and J. Zhang (2001): "The Two-Tier Labor Market in Urban China: Occupational Segregation and Wage Differentials Between Urban Residents and Rural Migrants in Shanghai," *Journal of Comparative Economics*, forthcoming.

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46(1), 69–85.

Schultz, T. P. (2001): "Productive Benefits of Improving Health: Evidence from Low-Income Countries," processed, Department of Economics, Yale University, March.

Schultz, T. P., and A. Tansel (1997): "Wage and Labor Supply Effects of Illness in Côte d'Ivoire and Ghana: Instrumental Variable Estimates for Days Disabled," *Journal of Development Economics*, 53(2), 251–286.

Semykina, A., and J. M. Wooldridge (2005): "Estimating Panel Data Models in the Presence of Endogeneity and Selection: Theory and Application," unpublished working paper, Michigan State University.

Steckel, R. H. (1995): "Stature and the Standard of Living," *Journal of Economic Literature*, 33(4), 1903–1940.

Strauss, J., and D. Thomas (1998): "Health, Nutrition and Economic Development," *Journal of Economic Literature*, 36(2), 766–817.

Thomas, D., and J. Strauss (1997): "Health and Wages: Evidence on Men and Women in Urban Brazil," *Journal of Econometrics*, 77(1), 159–185.

	Full sample			Subsample of wage earners		
Observations	3354			1216		
	mean	standard deviation		mean	standard deviation	
		total	"within" individual		total	"within" individual
hourly wage				2.2414	4.1085	2.8632
age	42.6956	13.3718	3.8316	40.6900	11.2849	3.6855
urban	0.3211	0.4670	0.0000	0.4918	0.5001	0.0000
public sector	0.3596	0.4799	0.1881	0.8026	0.3982	0.2304
state sector	0.2352	0.4242	0.2032	0.5362	0.4989	0.2953
Human capital						
height	1.6536	0.0672	0.0090	1.6736	0.0650	0.0081
years schooling	7.6839	3.8452	0.8188	9.7261	3.6943	0.8738
Instruments						
Ethnic one child policy	0.4213	0.4938	0.0429	0.3363	0.4726	0.0454
Newspaper availability	0.3560	0.4789	0.0431	0.4515	0.4978	0.0475

Table 1: Summary statistics

	1991	1993	1997	2000	2004
	(1)	(2)	(3)	(4)	(5)
age	− 0.1038	− 0.1326	−0.0417	− 0.0597	− 0.1278
	(0.0377)	(0.0319)	(0.0366)	(0.0463)	(0.0758)
	[−0.0352]	[−0.0494]	[−0.0142]	[−0.0201]	[−0.0432]
urban	0.5185	0.3580	− 0.2124	0.0692	0.0421
	(0.1707)	(0.1511)	(0.1814)	(0.2230)	(0.3082)
	[0.1818]	[0.1349]	[−0.0708]	[0.0236]	[0.0143]
public sector	1.4045	1.4470	1.3075	1.9587	0.3158
	(0.3789)	(0.2917)	(0.3337)	(0.3863)	(0.4446)
	[0.4743]	[0.5080]	[0.4708]	[0.6652]	[0.1119]
state sector	0.6850	−0.5539	0.3259	−0.3503	0.6974
	(0.3458)	(0.2707)	(0.4299)	(0.3646)	(0.6042)
	[0.2455]	[−0.1954]	[0.1165]	[−0.1104]	[0.2647]
Identification variables					
Ethnic one-child policy	0.4074	0.0870	− 0.2392	− 0.1375	0.0074
	(0.1508)	(0.1324)	(0.1630)	(0.1669)	(0.2332)
	[0.1403]	[0.0325]	[−0.0810]	[−0.0461]	[0.0025]
Newspaper availability	0.1145	− 0.1673	0.2880	0.2439	− 0.1025
	(0.1524)	(0.1519)	(0.1726)	(0.1920)	(0.2659)
	[0.0391]	[−0.0619]	[0.1008]	[0.0843]	[−0.0391]
observations	887	875	622	560	410

Table 2: First stage Mundlak probit estimations for 1991, 1993, 1997, 2000 and 2004. Standard errors in parenthesis. Marginal effects in square brakets.

	Wooldridge test	Semykina-Wooldridge	Semykina-Wooldridge
	(1)	(2)	(3)
age	0.1760 (0.0057)	−0.0036 (0.3459)	−0.0171 (0.1364)
urban		0.0031 (1.1570)	0.0161 (0.1676)
D93		0.2464 (1.2096)	0.2782 (0.3485)
D97		1.3388 (4.3896)	2.9363 (4.5276)
D00		1.6669 (2.4813)	3.2674 (4.2334)
D04		1.8307 (3.9925)	3.5266 (4.0955)
public sector	0.1657 (0.1100)	0.0162 (3.7675)	−0.1239 (0.6257)
state sector	−0.0217 (0.0606)	−0.2388 (0.9833)	−0.1876 (0.4209)
Human capital			
height / height 1991-1993	−1.7502 (7.0572)	4.7981 (20.9008)	2.2268 (6.9524)
height 1997-2004			1.0193 (7.1290)
years schooling / years schooling 1991-1993	0.0424 (0.0355)	0.0816 (0.0884)	0.0090 (0.1429)
years schooling 1997-2004			0.0639 (0.1849)
Inverse Mills ratio:			
$\widehat{\lambda}_{it}$	0.4893 (0.1058)	0.2932 (3.8592)	0.2224 (0.4649)
σ	0.5313		
R^2	0.0736	0.5732	0.6677
$F-$test: $\alpha_i = \alpha$	$F_{653}^{554}= 4.39$		
Hausman test	$\chi_8^2= 1844.37$		
Partial r^2 Height		0.0023	
Education		0.0280	
Partial F_test Height		$F_{1189}^4 = 0.40$	
Education		$F_{1189}^4 = 9.82$	
Hansen test of overidentification		$\chi_2^2= 0.3427$	$\chi_4^2= 0.6090$
observations	1216	1216	1216
individuals	555		
observations in first stage probit(s)	3354	3354	3354

Table 3: The returns to human capital in China, 1991, 1993, 1997, 2000 and 2004. Wooldridge test for the presence of sample selection bias, and Semykina-Wooldridge estimator (Huber-White standard errors in parentheses).