A Control Function Approach to Estimating the Returns to Schooling in Urban China*

Yafeng Wang†
Wang Yanan Institute for Studies in Economics
Xiamen University

Abstract

This paper applies a new estimator to estimate the returns to schooling in urban China. The identification via heteroskedasticity principle is used to control for the endogeneity of schooling when there are no valid instrumental variables. With the urban data in 2002 wave of Chinese Household Income Projects (CHIPs), we find, after controlling for the endogeneity of schooling, the estimates of returns to schooling over the sample period are far below the Ordinary Least Squares estimates which is in contrast to many alternative studies, which frequently find that the Ordinary Least Squares estimates are lower than the Instrumental Variables estimates. This lends to support the conventional wisdom that the endogeneity of schooling is due to the unobserved ability.

Keywords: Returns to Schooling, Control Function, Heteroskedasticity, Endogeneity.

JEL Classification Numbers: I2, J3.

*The author would like to thank participants at 2010 International Symposium on Contemporary Labor Economics for helpful discussions.

†Address: The Wang Yanan Institute for Studies in Economics (WISE), A301, Economics Building, Xiamen University, Xiamen, 361005, Fujian, China. E-mail: econyfwang@gmail.com

1 Introduction

Returns to schooling provide important information about the incentives for human capital accumulation, the efficiency of resource allocation, and the distributional consequences of differences in human capital, it may be the most commonly explored "treatment effect" in the empirical economics literature. Technically, the most concern about the estimation of returns to schooling is that the endogeneity of educational choices to wages, since it will bias the ordinary least squares (OLS) estimates. Strategies employed to account for the endogeneity of schooling are generally based on instrumental variables (IV) estimation (e.g., Angrist and Krueger, 1991; Card, 1999; Heckman *et al.*, 2006).

Since the endogeneity of educational choices to wages is typically attributed to the correlation between the unobservable factors (such as ability) which determine education levels and wages, a common feature of various IV approaches is that they exploit the existence of a variable(s) (IV) which is responsible for some variation in the conditional mean of the education level, but is exogenous and unrelated to wages. Unfortunately, such variables are not easily to exploit in the empirical research, especially when working with Chinese dataset, to the best of our knowledge, although there are numerous studies about the estimation of the returns to schooling in China[1], none of them considers the endogeneity problem[2].

An alternative strategy to the IV approach is to exploit the variation in the conditional error variances while imposing restrictions on other conditional second moments. In a seminal paper, Vella and Verbeek (1997) provide a rank order IV procedure, Rummery *et al.* (1999) use this strategy to estimate the returns to schooling for Australian

[1]Studies of the returns to schooling in China include Johnson and Chow (1997); Meng and Kidd (1997); Maurer-Fazio (1999); Li (2003); Zhang, Zhao *et al.* (2005); and Luo (2007).

[2]One exception is Liu (2007), for estimating the Chinese city-level external returns to education, he construted IV which is correlated with city average education and orthogonal to unobserved city-specific characteristics based on compulsory education law. In the individual level, there is no existing research which explored a valid IV to deal with the endogeneity problem in the estimation of Chinese returns to schooling.

youth. Although this rank order IV strategy provides an identifying source in the absence of exclusion restrictions, however, its value to empirical work is limited due to the nature of the error structures it allows. More recently, Klein and Vella (2010) provide an estimator for more general error structures that allow the heteroskedasticity in both equations to be functions of the same variables provided the correlation coefficient for the underlying homoskedastic error terms across equations is constant, the identification results are based on semiparametric representations of the heteroskedasticity. Although this semiparametric representation is theoretically appealing as the results are not reliant on specific parametric forms of the heteroskedasticity, the programming and computation requirements are demanding. Farré, Klein, and Vella (2010) (hereafter FKV) adapted the estimator proposed by Klein and Vella (2010) to a parametric setting thereby making it more easily to implement and illustrated the procedure by estimating the return to education using a sample of individuals from the 2004 wave of the National Longitudinal Survey of Youth.

In this paper, we employ the estimator proposed by FKV to estimate the returns to schooling in urban China but with several modifications. Firstly, the full parametric estimator proposed by FKV is obtained by multiple stages of nonlinear OLS estimations which is generally inefficient and the standard errors that result in the final stage are often incorrect, because they fail to account for estimation error in the previous stage(s). We modify this multi-stage estimator to a special case of generalized method of moments (GMM) estimator, efficient estimation and consistent standard errors are obtained by using the standard formulas for the efficient choice of weighting matrix. Secondly, we relax the heteroskedasticity form in the schooling equation to an unknown function of some exogenous covariates and use a full nonparametric estimation in this stage. Due to the feature of our data, the variables which affect the variance of schooling are all discrete, the nonparametric estimation with only discrete regressors is easy to implement and $\sqrt{n}$ consistent (frequency estimation). With the urban data in Chinese Household

Income Projects (CHIPs) conducted in 2002, we find, after accounting for the endogeneity of schooling, the estimates of returns to schooling are far below the OLS estimates, which is in contrast to many alternative studies which frequently find that the OLS estimate is lower than the IV estimate (e.g. Kling, 2001; Cameron and Taber, 2004), Ashenfelter *et al.* (1999) provide a survey of IV literature and show that the average difference between IV and OLS is around 3% per year of schooling[3]. The estimates and relative difference between our estimates and OLS estimates vary with different set of control variables (0.046 vs. 0.073; 0.028 vs. 0.046; 0.023 vs. 0.039). As a robustness test, we use Manski and Pepper's (2000) nonparametric identification strategy to estimate the upper bounds on the returns to schooling, the upper bounds are below OLS estimates which is consistent with our main findings. Our results suggest that the OLS estimate of returns to schooling in urban china are biased upwards, the conventional wisdom that the endogeneity of educational choices to wages is typically attributed to the correlation between the unobservable factors (such as ability) which determine education levels and wages is confirmed.

The paper is organized as follows. In the next section we discuss the FKV's identification strategy in the returns to schooling context, our modifications are also presented. We also outline how such procedures can be implemented. In section 3 we describe our data, present descriptive statistics for our sample. Estimates of the returns to schooling in urban China are presented in Section 4. Section 5 concludes.

[3] Ashenfelter *et al.* (1999) also find that the average premium of around 3% of IV over OLS may be partly (1.8%) explained by selective reporting of results by researchers.

2 The Model

2.1 Model and Identification

We begin by representing FKV's approach for identification and estimation of the returns to schooling. Although we make several modifications in the implementation, we do not provide any new theoretical results, and this section is draw heavily from FKV. The model contains two equations and has the following triangular form:

$$\ln w_i = x_i'\beta_0 + schooling_i\beta_1 + u_i \tag{1}$$

$$schooling_i = x_i'\delta_0 + v_i \tag{2}$$

where (1) is a semi-logarithmic specification for earnings known as extended Mincer equation and (2) is the schooling equation. w_i is earnings of individual i, $schooling_i$ is educational attainment measured as years of schooling, and x_i denotes a vector of exogenous variables such that:

$$E[u_i|x_i] = E[v_i|x_i] = 0 \tag{3}$$

The endogeneity of *schooling* arises through the possible correlation between the error terms u_i and v_i which renders the OLS estimates of the β's inconsistent[4]. Following FKV, we specify the same x_i appear in (1) and (2) to reveal there are no available instruments to estimate (1), actually, the variables which appear in wage equation but not appear in schooling equation do not identify the model as IV requires variable(s) in the schooling equation which do not appear in the wage equation.

To identify the model without any valid IV, FKV assume the presence of heteroskedasticity and impose an additional restriction on the correlation between the error terms. More explicitly, let $S_u^2(x_i)$ and $S_v^2(x_i)$ denote the conditional variance functions for u_i and

[4] u_i and v_i may be the correlated measures of unobserved abilities.

v_i and assume:

$$u_i = S_u(x_i)u_i^* \text{ and } v_i = S_v(x_i)v_i^* \tag{4}$$

where u_i^* and v_i^* are homoskedastic error terms. The key identifying restrictions contain two assumptions:

Assumption 1 *Either $S_u(x_i)$ and/or $S_v(x_i)$ are not constant and the ratio $S_u(x_i)/S_v(x_i)$ is not constant across i.*

Assumption 2 *The conditional correlation coefficient between the homoskedastic error terms, u_i^* and v_i^* is constant:*

$$E[u_i^* v_i^*] = E[u_i^* v_i^* | x_i] = \rho_0 \tag{5}$$

Assumption 1 states that at least one of the error terms is heterogenous, which suggests that the contribution of unobserved factors (such as ability) to wages and schooling depends on the individual's socioeconomic factors which is a reasonable restriction with economic implications since the presence of heteroskedasticity is largely an empirical issue. Also this restriction can be tested via White or other heteroskedasticity tests. One may argue the reality of assumption 2 since it requires that after conditioning out the role of the x_i the return to unobserved ability is constant, Klein and Vella (2010) show that this constant conditional correlation assumption is consistent with a number of data generating processes. Based on the fact that the distributions of the error terms does depend on x_i, Klein and Vella (2010) and FKV show that the identification can be achieved by making a new error term in (1) conditional on x_i:

$$\epsilon_i = u_i - A(x_i)v_i \tag{6}$$

where

$$A(x_i) = \rho_0 \frac{S_u(x_i)}{S_v(x_i)}$$

and $A(x_i)$ is a nonlinear function of x_i, this non linearity in $A(x_i)$ is a source of identification provided one can impose the appropriate structure in estimation. Substitute (6) into (1) then we can consistently estimate β from the following controlled regression:

$$\ln w_i = x_i'\beta_0 + schooling_i\beta_1 + \rho_0 \frac{S_u(x_i)}{S_v(x_i)} v_i + \epsilon_i \tag{7}$$

where ϵ_i is a zero mean error term, and does not correlate with regressors. The main difficulty in the estimation of (7) comes from the estimation of $S_u(x_i)$, since both v_i and $S_v(x_i)$ are straightforward to estimate given the specification (2).

2.2 Estimation

To estimate the controlled regression (7), we need to specify the mechanism that the x_i enter the $S_u(x_i)$ and $S_v(x_i)$ functions. Klein and Vella (2010) impose single index structure for the conditional variance functions:

$$S_u^2(x_i) = E[u_i^2 | I_u(\theta_u)] \text{ and } S_v^2(x_i) = E[v_i^2 | I_v(\theta_v)] \tag{8}$$

and use (iterated) semiparametric least squares to estimate the underlying parameters. The computational demands associated with this procedure arise in the estimation of the main equation since the semiparatetric nature of the $S_u(x_i)$ function and the non linearity inherent in estimating the parameters in (7) requires estimating this semiparametric function multiple times in each round of each iteration of the optimization problem. Due to the computational difficulties associated with estimating these functions, FKV suggest estimate the model by treating both $S_u(x_i)$ and $S_v(x_i)$ as known functions of an index with unknown parameters:

$$S_u^2(x_i) = \exp(z'_{ui}\theta_u) \text{ and } S_v^2(x_i) = \exp(z'_{vi}\theta_v) \tag{9}$$

where z_{ui} and z_{vi} are the vector of variables considered to produce the heteroskedasticity in the respective equations and they can be differ from the x_i. Such specification of heteroskedasticity forms can be tested by using different parameterizations of the heteroskedasticity.

Based on the parameterization of heteroskedasticity in (9), FKV propose to estimate β via iterated nonlinear OLS, which is a sequential estimator. Although FKV suggest to employ one additional step to separates the estimation of the β from the estimation of S_u, like the usual two stage estimations, this estimation procedure is generally inefficient, and the standard errors that result in the final stage are often incorrect, because they fail to account for estimation error in the first stage (Newey, 1984). Since all the stages in FKV's estimation procedure are nonlinear least square estimations, which can be easily modified to a standard GMM estimation framework, we propose to estimate the model parameters through a GMM estimation. First define the moment conditions corresponding to the least square problems of FKV's estimation procedure, let $Y = (\ln w, x', shooling, z'_u, z'_v)'$ and

$$g(Y, \beta, \theta, \rho_0, \delta) = \begin{bmatrix} g_1(Y, \delta) \\ g_2(Y, \theta_v, \delta) \\ g_3(Y, \beta, \theta_u, \delta) \\ g_4(Y, \beta, \theta, \rho_0, \delta) \end{bmatrix} \tag{10}$$

where

$$g_1(Y, \delta) = [schooling - x'\delta_0]x$$

$$g_2(Y, \theta_v, \delta) = [\ln(schooling - x'\delta_0)^2 - z_v'\theta_v]z_v$$

$$g_3(Y, \beta, \theta_u, \delta) = [\ln(\ln w - x'\beta_0 - schooling\beta_1)^2 - z_u'\theta_u]z_u$$

$$g_4(Y, \beta, \theta, \rho_0, \delta) = [\ln w - x'\beta_0 - schooling\beta_1 - \rho_0\sqrt{\exp(z_u'\theta_u)}\frac{[schooling - x'\delta_0]}{\sqrt{\exp(z_v'\theta_v)}}]$$

$$\begin{bmatrix} x \\ schooling \\ \frac{\sqrt{\exp(z_u'\theta_u)}[schooling - x'\delta_0]}{\sqrt{\exp(z_v'\theta_v)}} \end{bmatrix}$$

Then from the model assumptions, it is clearly that

$$E[g(Y, \beta, \theta, \rho_0, \delta)] = 0 \tag{11}$$

Thus the joint GMM estimator takes the form

$$\hat{\beta}, \hat{\theta}, \hat{\rho}_0, \hat{\delta} = \arg \min_{\{\beta,\theta,\rho_0,\delta\}} \sum_{i=1}^{n} g(Y_i, \beta, \theta, \rho_0, \delta)'\Omega_n \sum_{i=1}^{n} g(Y_i, \beta, \theta, \rho_0, \delta) \tag{12}$$

We refer to this estimator as **FGMM**, since here we use a fully parametric specification. Efficient estimation and consistent standard errors are obtained by using the standard formulas for the efficient choice of Ω_n. Given currently existing computing power and readily available automated GMM estimation programs, this general procedure should be broadly applicable, especially since the iterated OLS estimator can themselves provides good, consistent starting values for estimation. Also, to minimize or avoid numerical searches if necessary, asymptotic efficiency can be obtained without iterating to convergence. Newey and McFadden (1986) show that asymptotic efficiency is obtained by just doing one iteration of the efficient GMM estimator.

To avoid the potential misspecification of the conditional variance function $S_v^2(x_i)$, we also propose to use the nonparametric specification instead of the parametric one for $S_v^2(x_i)$:

$$S_u^2(x_i) = \exp(z'_{ui}\theta_u) \text{ and } S_v^2(x_i) = E[v_i^2|z_{vi}] \tag{13}$$

Although FKV also use a semiparametric estimator for S_v^2, our use of nonparametric estimation is differ from their setting which is due to the special feature of our dataset and empirical model setting, which may not be general as their framework. A interesting feature of our empirical model is that variables in z_v are all discrete, which makes the nonparametric estimation of $E[v_i^2|z_{vi}]$ be simple and without the usual curse of dimension of the nonparametric estimation. The frequency estimator for $E[v_i^2|z_{vi}]$ when z_v contains only discrete variables takes the form:

$$\widehat{E[v_i^2|z_{vi}]} = \frac{1}{n}\sum_{i=1}^{n} v_i^2 I(z_{vi} = z_v)/\tilde{p}(z_v) \tag{14}$$

where

$$\tilde{p}(z_v) = \frac{1}{n}\sum_{i=1}^{n} I(z_{vi} = z_v) \tag{15}$$

is the estimated probability function. With this $\sqrt{n}$ consistent $\widehat{E[v_i^2|z_{vi}]}$ we can use the following iterated estimation procedure to get the final estimates of the underlying parameters:

1. Regress $schooling_i$ on x_i to obtain a consistent estimate of the residual which is denoted by $\hat{v}_i$.
2. Estimate $\hat{S}_v$ through frequency estimation (14), and $\hat{S}_v = \sqrt{\widehat{E[\hat{v}_i^2|z_{vi}]}}$.
3. Using $\hat{v}_i$ and $\hat{S}_v$ estimate the model parameters through a iterative procedure. For a given value of β, say β_c, define the residual $u_i(\beta_c)$. Using this value of $u_i(\beta_c)$, regress $\ln(u_i^2(\beta_c))$ on z_{ui} get θ_{cu} and compute $\hat{S}_{ui}(\beta_c)$ as $\sqrt{\exp(z'_{vi}\theta_{cu})}$. Then estimate ρ_{0c}

as:

$$\min_{\rho_{0c}} \sum_{i=1}^{n} \left(u_i(\beta_c) - \rho_{0c} \hat{S}_{ui}(\beta_c) \frac{\hat{v}_i}{\hat{S}_v(x_i)} \right)^2 \tag{16}$$

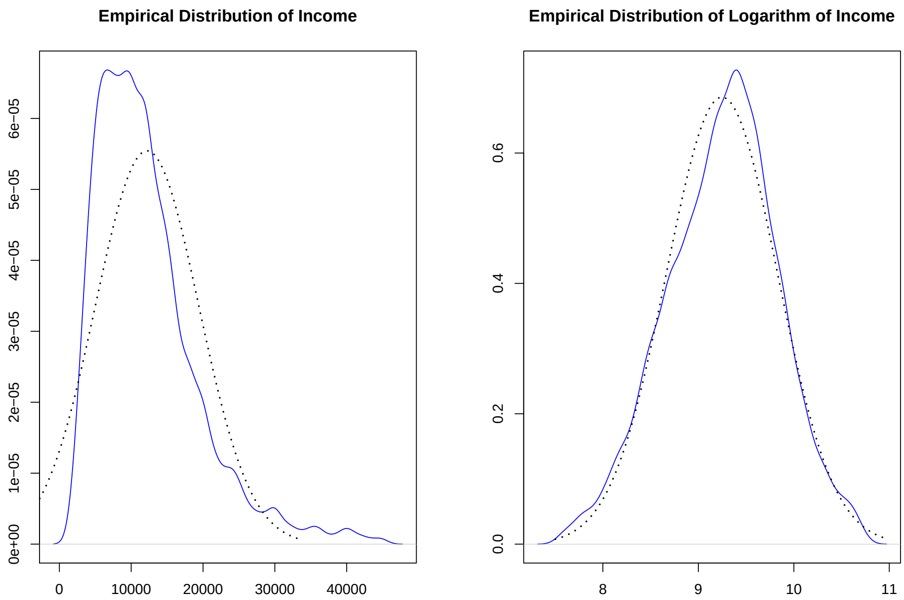
The final estimates of β_c, ρ_{0c}, and θ_{cu} are those that minimize (16) and are obtained through a standard iterative procedure. The consistent standard errors are obtained via bootstrap.

We refer to this estimator as **INPOLS**.

3 Data

The data used in this paper is the survey of urban households which comes from 2002 wave of Chinese Household Income Projects (CHIPs) conducted by Chinese Academy of Social Sciences (CASS). This survey contains 20632 persons from 6835 households in 12 provinces that are broadly representative of China's rich regional variation. To focus on wage determination in the labor market, we restrict our sample to workers engaged in wage employment. Following standard practice, we exclude employers, self-employed individuals, retirees, students, and household workers (e.g. Coleman, 1993; Mwabu and Schultz, 1996). Moreover, as China's Labor Law sets the minimum working age at 16, we exclude all those younger than 16. Because most workers retire by age 60 in accordance with China's mandatory retirement age, individuals older than 60 also are excluded. Wage income consists of four major components, namely, basic wage, bonus, subsidies and other labor-related income, Figure 1 reports the empirical distribution of wage and logarithmic value of wage. The major control variables in the wage equation are experience (*exper*) which is a worker's years of potential labor market experience measured as age minus schooling minus six, the square of *exper*, gender (*gender*) which is a dummy variable capturing the wage differential between men and women, marriage (*marriage*) which is also a dummy variable capturing the effect of with-spouse, and party (*party*) which is

Figure 1: Empirical Distribution of Income



a dummy variable capturing the wage differential between any party membership and non-party. Table 1 reports the descriptive statistics for these variables.

One desirable feature of 2002 CHIPs data is that it contains detailed information for the head of household and her spouse's parents, which makes us can control the parents' features in the estimation of the returns to schooling. Moreover, the parents' characteristics also have impact on years of education attainment, we control these variables in both wage and schooling equations, while containing these variables means that we should match the individual with her parents, which restricts our sample size to 8453. Beyond the parents' characteristics, we also consider controlling the employment characteristics, which includes type of employment, ownership, professional nature and industry. Another set of control variables is the Region which is a set of 11 provincial dummy variables. A detailed definition of these control variables is reported in Table 2.

Table 1: Descriptive Statistics

Variable	N	Mean	Std. Dev.	Min	Max
income	8453	12210.780	7196.130	1947	44999.980
lincome	8453	9.247	0.582	7.574	10.714
schooling	8453	11.299	2.989	0	23
exper	8453	24.763	9.106	−5	50
gender	8453	0.556	0.497	0	1
marriage	8453	0.951	0.216	0	1
party	8453	0.330	0.470	0	1

4 Estimation Results

For estimating the returns to schooling in urban China using the procedures described above, we use the specific model:

$$\ln w_i = x'_{wi}\beta_0 + schooling_i\beta_1 + u_i \tag{17}$$

$$schooling_i = x'_{si}\delta_0 + v_i \tag{18}$$

where $x_{si} \subseteq x_{wi}$ indicates that we have no valid instrumental variables for the endogenous variable *schooling*. The existing researches using CHIPs data all ignore the endogenous problem, here we use the control function approach to deal with this problem and compare the difference with the usual OLS estimates. The explanatory variables are those commonly employed in the estimation of returns to schooling which capture the individual's background, parents' characteristics and employment characteristics. For the schooling equation, we set x_{si} be the parents' characteristics, which includes parents' party membership, education level and professional nature. While for the wage equation, we consider three specifications for x_{wi}, the common variables among them are the experience, square of experience, individual's party membership, gender and individual's marriage status, the difference among them is that the first specification also consider the parents' char-

Table 2: Definition of Control Variables

Employment Characteristics			
Type	Ownership	Professional Nature	Industry
$wa1$ I(profit enterprises)	$ow1$ I(state-owned)	$wpro1$ I(person in charge)	$industry1$
$wa2$ I(loss-making enterprises)	$ow2$ I(collective)	$wpro2$ I(technical staff)	:
$wa3$ I(government departments)	$ow3$ I(private)	$wpro3$ I(skilled workers)	$industry15$
$wa4$ I(institutions)	$ow4$ I(foreign*)		

Parents Characteristics					
Education Level		Professional Nature		Party	
Farther	Mother	Farther	Mother	Farther	Mother
$feduc1$	$meduc1$	$fwpro1$	$mwpro1$	$fparty$	$mparty$
I(university and above)		I(person in charge)		I(party**)	
$feduc2$	$meduc2$	$fwpro2$	$mwpro2$		
I(college)		I(technical staff)			
$feduc3$	$meduc3$	$fwpro3$	$mwpro3$		
I(secondary)		I(skilled workers)			
$feduc4$	$meduc4$				
I(high School)					
$feduc5$	$meduc5$				
I(Junior)					

*Foreign includes the joint venture; **party includes communist and other parties acteristics, the second specification considers parents' characteristics and employment characteristics, while the last specification adds the region dummy variables. For different sets of control variables, we report the OLS[5], FGMM and INPOLS estimates. Note that although the GMM estimation only contains one step, we report the estimates for different equations separately.

4.1 Schooling Equation

We first discuss the FGMM estimates of the schooling equation which are reported in the second column of Table 3. The estimates are consistent with those in the existing educational literature. Parental education and mother's professional nature have an important positive effect on years of education, while the impact of father's professional nature is not significant.

The control function approach introduced above requires at least one of the equations' error terms to be heteroskedastic. Using the estimates from Table 3 we examine the presence of heteroskedasticity of the schooling equation. The statistic for the White test is 219 and that for the Breusch-Pagan, using all the explanatory variables in the model, is 162.93. These values clearly reject the null hypothesis of homoskedastic errors.

The next step is to determine the form of heteroskedasticity in the schooling equation, S_{vi}^2. An examination of the results of the heteroskedasticity tests suggests that the main variables responsible for the heteroskedasticity are gender and mother's educational level. Though we suspect that some of the variables in the schooling model may not affect the error variance we do not have strong arguments to exclude them from the heteroskedastic index, since the joint significance can not be rejected. Accordingly in estimating the conditional variance function for the schooling equation we use all variables which appeared in the conditional mean (i.e., $z_{vi} = x_{si}$).

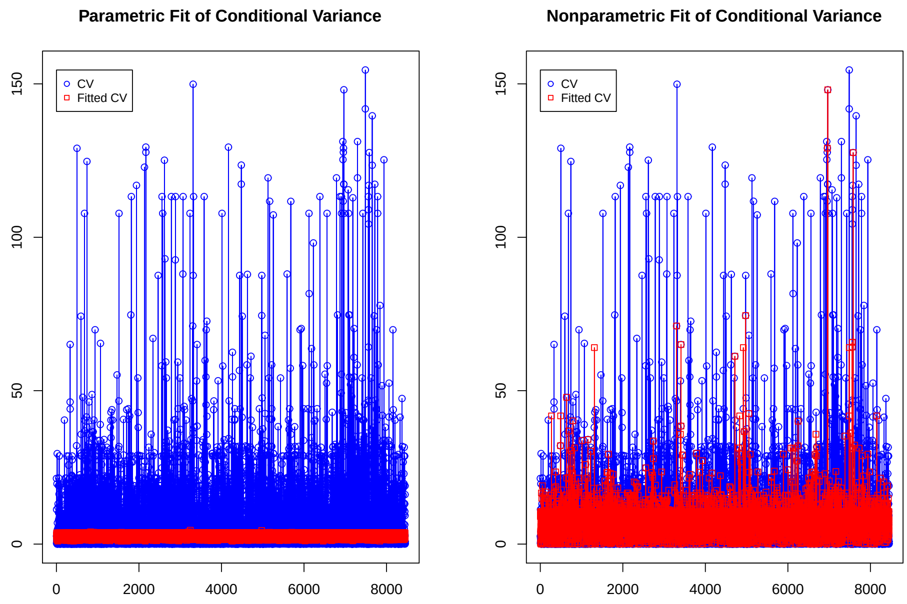
---

[5]For schooling equation we only use OLS to do the usual test of the presence of heteroskedasticity.

Table 3: Parametric Estimation for Schooling Equation

Variable	Conditional Mean		Conditional Variance	
	estimates	SD	estimates	SD
sex	0.2604	(0.0592)	0.2002	(0.0433)
fparty	0.1886	(0.0623)	−0.1223	(0.0442)
mparty	0.0617	(0.0589)	−0.0179	(0.0456)
*feduc*1	1.1160	(0.1773)	−0.1674	(0.1302)
*feduc*2	1.4667	(0.1727)	−0.1691	(0.1316)
*feduc*3	1.1902	(0.1489)	−0.1032	(0.1097)
*feduc*4	0.8116	(0.1206)	−0.3120	(0.0848)
*feduc*5	0.5430	(0.0848)	−0.1567	(0.0589)
*meduc*1	1.1869	(0.3281)	−0.2521	(0.0627)
*meduc*2	0.6731	(0.2731)	−0.2544	(0.0938)
*meduc*3	0.7274	(0.1914)	−0.1931	(0.0622)
*meduc*4	0.7877	(0.1483)	0.3204	(0.2363)
*meduc*5	0.5273	(0.0980)	−0.0416	(0.2049)
*fwpro*1	0.1868	(0.0918)	0.1856	(0.1301)
*fwpro*2	0.1893	(0.1345)	0.0475	(0.1080)
*fwpro*3	−0.0182	(0.0898)	−0.0825	(0.0741)
*mwpro*1	0.5056	(0.1142)	−0.1396	(0.0892)
*mwpro*2	0.6163	(0.1640)	0.0215	(0.1148)
*mwpro*3	0.1764	(0.1064)	−0.1477	(0.0774)
constant	10.3819	(0.0659)	1.1036	(0.0441)
Test for Heteroskedasticity (Based on OLS)			statistics	
White			219	
Breush-Pagan			162.93	

Figure 2: Parametric and Nonparametric Fits of Conditional Variance



The nonlinear least squares estimates of S_{vi}^2 are reported in the third column of Table 3. Moreover, the standard errors which come from the efficient GMM estimation are not suffer from the usual multi-stage problem. We also consider the nonparametric estimation of the conditional variance S_{vi}^2, the nonparametric fit of S_{vi}^2 is reported in Figure 2, and for comparison, we also report the parametric fit of this conditional variance in Figure 2. The mean square error (MSE) for the nonparametric fit is 169 while for the parametric fit is 223, which means that the we should use a nonparametric version instead of the parametric one for the conditional variance. But as can be found in the final estimates for the returns to schooling in the wage equation, the impact of the choice of the specification for conditional variance is not important if the number of control variables is rich enough, this also suggests FKV's parametric model.

4.2 Wage Equation

We now turn to the estimation of the primary equation. As noted above, we consider three sets of control variables for the wage equation, the common variables among them are the experience, square of experience, individual's party membership, gender and individual's marriage status. In addition to these variables, the first specification includes parents' characteristics, the second specification includes parents' characteristics and employment characteristics, the last specification considers all these features but adds the region dummy variables.

Before considering the control function estimates we briefly discuss the OLS estimates reported in the second columns of Table 4 to 6. The primary feature of interest is the estimated impact of education on earnings which are 0.071, 0.0459 and 0.0386 for three specifications respectively. The magnitude of these coefficients are in line with the previously reported OLS estimates in Luo (2007) which use the same data.

In implementing our estimation strategy it is first necessary to specify the variables entering the heteroskedasticity index of the wage equation, z_{ui}. Although we experimented with different choices for the variables, the results reported here are based on all the variables in x_{wi} and a constant (i.e. $z_{ui} = x_{wi}|1$). The third columns of Table 4 to 6 present the estimates of the coefficients in the wage equation obtained using the method which we refer to as FGMM in section 2.2. And the fourth columns of Table 4 to 6 present the estimates of the coefficients in the wage equation obtained using the method which we refer to as INPOLS in section 2.2.

Before focusing on the estimated impact of schooling on wages we highlight a number of the interesting features of this table. First, the OLS, FGMM and INPOLS estimates for the exogenous variables are generally quite similar. All these estimates provide evidence of a small marriage premium and a gender differential. The impact of party membership is positive and significant which states the importance of political identity in China. The

Table 4: Estimation for Wage Equation

(Only control the Parents Characteristics)			
	OLS	FGMM	INPOLS
schooling	0.0731	0.0420	0.0463
	(0.0025)	(0.0121)	(0.0076)
exper	0.0179	0.0181	0.0172
	(0.0032)	(0.0032)	(0.0032)
*exper*2	-0.0001	-0.0001	-0.0001
	$(6.22e-05)$	$(6.38e-05)$	$(6.32e-05)$
gender	0.1675	0.1764	0.1751
	(0.0115)	(0.0126)	(0.0120)
marriage	0.0260	0.0269	0.0288
	(0.0333)	(0.0335)	(0.0333)
party	0.1211	0.1243	0.1225
	(0.0128)	(0.0136)	(0.0135)
ρ_0		0.1794	0.2878
		(0.0672)	(0.0775)
constant	7.8744	8.1976	8.1613
	(0.0590)	(0.1411)	(0.1010)
Test for Heteroskedasticity	(Based on OLS)	Statistics	
White		358.32	
Breush-Pagan		96.61	

joint significance test for different sets of control variables in different settings all suggest the significance of these variables, and when the number of control variables increases, the impacts of all the control variables decrease. Finally, the test of the presence of heteroskedasticity based on the OLS estimation suggests the existence of heteroskedasticity in the wage equation.

The key feature our results reported in Table 4 to 6, however, is the difference among the estimates of the returns to schooling. While the OLS estimates were 0.071, 0.0459 and 0.0386, when accounting for the endogeneity of schooling, the FGMM estimates

Table 5: Estimation for Wage Equation

(Control the Parents and Employment Characteristics)			
	OLS	FGMM	INPOLS
schooling	0.0459	0.0278	0.0280
	(0.0025)	(0.0109)	(0.0066)
exper	0.0234	0.0241	0.0235
	(0.0030)	(0.0029)	(0.0029)
*exper*2	-0.0003	-0.0003	-0.0003
	$(6.02e-05)$	$(5.79e-05)$	$(5.77e-05)$
gender	0.1376	0.1431	0.1429
	(0.0122)	(0.0123)	(0.0117)
marriage	0.0085	0.0089	0.0102
	(0.0320)	(0.0310)	(0.0310)
party	0.0667	0.0678	0.0673
	(0.0121)	(0.0129)	(0.0129)
ρ_0		0.1171	0.2157
		(0.0682)	(0.0758)
constant	8.0311	8.2151	8.2186
	(0.0669)	(0.1333)	(0.0990)
Test for Heteroskedasticity	(Based on OLS)	Statistics	
White		–	
Breush-Pagan		271.66	

Table 6: Estimation for Wage Equation

(Control the Parents, Employment Characteristics and Regions)			
	OLS	FGMM	INPOLS
schooling	0.0386	0.0234	0.0232
	(0.0024)	(0.0098)	(0.0062)
exper	0.0186	0.0190	0.0185
	(0.0029)	(0.0028)	(0.0028)
*exper*2	−0.0002	−0.0003	−0.0002
	$(5.7e-05)$	$(5.47e-05)$	$(5.45e-05)$
gender	0.1567	0.1612	0.1610
	(0.0104)	(0.0111)	(0.0107)
marriage	0.0583	0.0589	0.0603
	(0.0291)	(0.0286)	(0.0286)
party	0.0756	0.0767	0.0763
	(0.0118)	(0.0117)	(0.0115)
ρ_0		0.1081	0.2032
		(0.0674)	(0.0782)
constant	7.9472	8.1031	8.1095
	(0.0689)	(0.1218)	(0.0928)
Test for Heteroskedasticity		(Based on OLS)	Statistics
White			–
Breush-Pagan			85.34

are 0.042, 0.0278, and 0.0234, the INPOLS estimates are 0.0463, 0.0280 and 0.0232. Moreover while there is some loss in statistical significance, in comparison to the OLS estimate, the FGMM and INPOLS estimates are statistically significant at conventional levels of testing. Finally the estimate of the correlation coefficient, ρ_0, is positive and statistically significant, indicating that schooling is not exogenous. Our results suggest, after accounting for the endogeneity of schooling, the estimates of returns to schooling are far below the OLS estimates which is in contrast to many alternative studies which frequently find that the OLS estimate is lower than the IV estimate. Our results are consistent with the conventional wisdom that the OLS estimates of returns to schooling are biased upwards, and the endogeneity of educational choices to wages is typically attributed to the correlation between the unobservable factors (such as ability) which determine education levels and wages.

The estimates and relative difference between our estimates and OLS estimates vary with different sets of control variables, and when the control variables are rich enough, the difference between FGMM and INPOLS estimates is quite small which suggests that the parametric model of the conditional variance is acceptable in the empirical research, while the INPOLS procedure can be used as the robust check or test.

4.3 Robust Analysis

Although our empirical results confirm the conventional wisdom that the OLS estimates of returns to schooling are biased upwards, many of the IV studies show that returns are often considerably larger than those found by OLS (e.g. Kling, 2001; Cameron and Taber, 2004). As a robust test, we make use of the nonparametric identification strategy proposed Manski and Pepper (2000) to estimate the nonparametric upper bounds on the average treatment effect of one year of education on earnings (the returns to schooling). The logic is that the true return should not be larger than the nonparametric upper bound, and if the OLS estimates are larger than the estimated upper bounds, we can conclude that the OLS estimates are biased upwards.

The average treatment effect of one year of education on earnings can be defined as

$$\Gamma(s,t) = \frac{\Delta(s,t)}{t-s} \tag{19}$$

where

$$\Delta(s,t) = E[y(t)] - E[y(s)]$$

for $s < t$. s and t are years of education and $y(\cdot)$ is log-earnings. Under assumptions of monotone treatment response and monotone treatment selection[6], Manski and Pepper (2000) derive the upper bound on $\Delta(s,t)$ as

$$\begin{aligned} \overline{\Delta}(s,t) = \quad & \sum_{u<s} (E[y(t)|z=t] - E[y(t)|z=u]) \cdot \Pr[z=u] \\ & +(E[y(t)|z=t] - E[y(t)|z=u]) \cdot \Pr[s \leq z \leq t] \\ & +\sum_{u>t} (E[y(t)|z=u] - E[y(t)|z=s]) \cdot \Pr[z=u] \end{aligned} \tag{20}$$

Then the upper bound on the returns to schooling will be

$$\bar{\Gamma}(s,t) = \frac{\bar{\Delta}(s,t)}{t-s} \tag{21}$$

Table 7 and Figure 3 report the results of such analysis of our data. Comparing with 6 years of schooling, the estimated upper bounds are below the OLS estimates except for 7 and 8 years of schooling, which suggests that in urban China, the OLS estimates of returns to schooling are biased upwards and when accounting for the endogenous bias,

[6]The monotone treatment response assumption states that for any $t_2 \geq t_1$, there is $y(t_2) \geq y(t_1)$, while the monotone treatment selection assumption states that for any $u_2 \geq u_1$, there is $E[y(t)|z = u_2] \geq E[y(t)|z = u_1]$, where z is the realized years of schooling.

Table 7: Nonparametric Upper Bounds on Returns to schooling

s	t	Upper Bound	s	t	Upper Bound
6	7	0.2129	6	15	0.0515
6	8	0.1039	6	16	0.0582
6	9	0.0770	6	17	0.0480
6	10	0.0684	6	18	0.0477
6	11	0.0531	6	19	0.0387
6	12	0.0494	6	20	0.0594
6	13	0.0575	6	21	0.0350
6	14	0.0517	6	22	0.0435
OLS			0.0886		

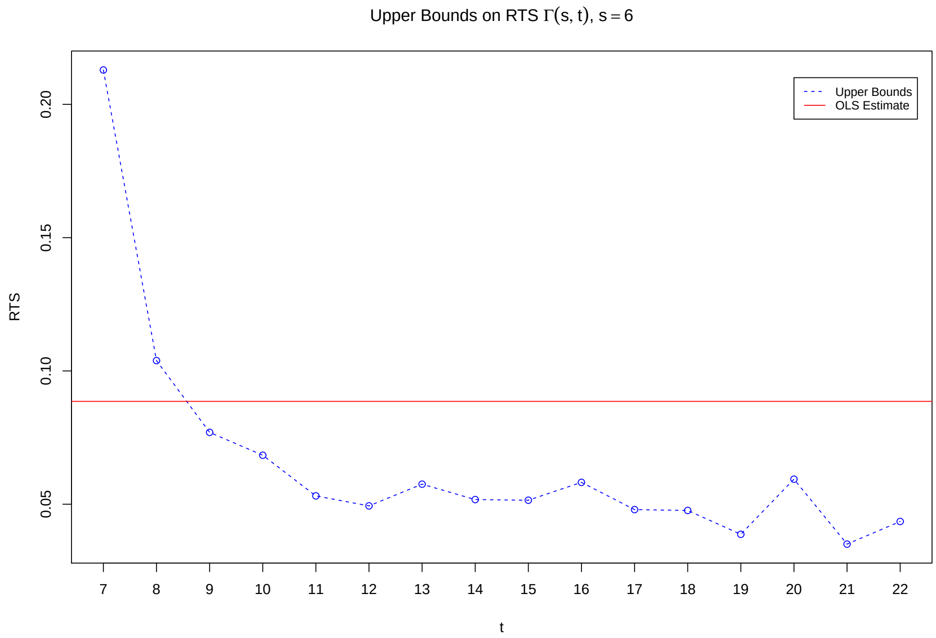
the estimates should be below the OLS estimates, that is our main results based on the control function approach.

5 Conclusion

In this paper, we employ a new estimator to estimate the returns to schooling in urban China. The endogeneity of schooling is controlled using only the natural heteroscedasticity of earnings and schoolings. Based on the full parametric framework proposed by FKV, which is to exploit the dependence of the errors on exogenous variables (e.g. heteroscedasticity), we propose two estimators. One is the GMM estimator using the same specification as FKV but take accout of the problem of multi-stage (sequential) estimators, efficient estimation and consistent standard errors are obtained by using the standard formulas for the efficient choice of weighting matrix. Another estimator which we refer to as INPOLS relaxes the heteroskedasticity form of schooling to an unknown function of exogenous variables. Unlike the usual IV approach, this control function method does not rely on exclusion restrictions.

We apply the method to the urban data in 2002 wave of Chinese Household Income

Figure 3: Nonparametric Upper Bounds on Returns to schooling



Projects (CHIPs) and find that the estimates of returns to schooling are 0.0463, 0.0280 and 0.0232 (depend on different sets of control variables). These estimates are far below the OLS estimates, which are 0.071, 0.0459 and 0.0386 correspondingly. Our result means that the OLS estimates are biased upwards which is in contrast with many alternative studies which frequently find that the OLS estimate is lower than the IV estimate. This lends to support the conventional wisdom that the endogeneity of schooling is due to the unobserved ability, since the estimated correlation coefficients between errors are significiantly positive. Also, the small difference between the full parametric estimation and the one which includes nonparametric part indicates that the full parametric framework proposed by FKV is acceptable in the empirical research. We check the robustness of our estimation by applying the nonparametric identification strategy proposed by Manski and Pepper (2000) to the same dataset, the result that OLS estimates are larger than the upper bounds is consistent with our main findings.

References

Angrist, Joshua D. and A. B. Krueger, 1991. Does compulsory school attendance affect schooling and earnings. *The Quarterly Journal of Economics* 106, 979-1014.

Ashenfelter, Orley, C. Harmon and H. Oosterbeek, 1999. A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics* 6, 453-470.

Cameron, S. V. and C. Taber, 2004. Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy* 112, 132-182.

Card, D., 1999. The causal effect of education on earnings. In Handbook of Labor Economics. Edited by Ashenfelter and D. Card. Amsterdam: Elsevier.

Coleman, M. T., 1993. Movements in the earnings-schooling relationship, 1940-88. *Journal of Human Resources* 28, 660-680.

Farré, L., R. Klein and F. Vella, 2010. A parametric control function approach to estimating the returns to schooling in the absence of exclusion restrictions: an application to the NLSY. IZA Discussion Paper No. 4935.

Heckman, J. J., S. Urzua and E. Vytlacil, 2006. Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics* 88, 389-432.

Klein, R. and F. Vella, 2010. Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics* 154, 154-164.

Kling, J. R., 2001. Interpreting instrumental variables estimates of the returns to schooling. *Journal of Business & Economic Statistics* 19, 358-364.

Liu, Z., 2007. The external returns to education: evidence from Chinese cities. *Journal of Urban Economics* 61, 542-564.

Luo, C., 2007. Return to education and Its distribution in urban China. *Economic Research Journal* 6, 119-129. (in Chinese)

Manski, C. F. and J. V. Pepper, 2000. Monotone instrumental variables with an application to the returns to schooling. *Econometrica* 68, 997-1012.

Mwabu, G. and T. P. Schultz, 1996. Education returns across quantiles of the wage function: alternative explanations for returns to education by race in south Africa. *American Economic Review* 86, 335-339.

Newey, W. K., 1984. A method of moments interpretation of sequential estimators. *Economics Letters* 14, 201-206.

Newey, W. K. and D. McFadden, 1986. Large sample estimation and hypothesis testing. In Handbook of Econometrics. Edited by Engle, R. F., and D. McFadden. Amsterdam: Elsevier.

Rummery, S., F. Vella and M. Verbeek, 1999. Estimating the returns to education for Australian youth via rank-order instrumental variables. *Labour Economics* 6, 491-507.

Vella, F. and M. Verbeek, 1997. Using Rank Order as an Instrumental Variable: An Application to the Return to Schooling. CES Discussion Paper No. 10.