Economic Modelling 47 (2015) 166–173

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/ecmod

Estimating returns to schooling in urban China using conventional and heteroskedasticity-based instruments☆

Vinod Mishra *, Russell Smyth [1]

Department of Economics Monash University, VIC 3800, Australia

ARTICLE INFO

Article history:
Accepted 1 February 2015
Available online xxxx

Keywords:
China
Schooling
Wages

ABSTRACT

A common problem when estimating the returns to schooling is the lack of conventional instrumental variables (IV) for education altogether or, if conventional IVs exist, there is often doubt as to whether they satisfy the exclusion restriction. We illustrate how a novel identification strategy, proposed by Lewbel (2012), which utilizes a heteroscedastic covariance restriction to construct an internal IV, can be used to obtain IV estimates for returns to education when alternatively there are no conventional IVs available or the conventional IVs that are available are potentially weak. To do so, we estimate the returns to schooling in urban China using two datasets. One dataset, which is for Shanghai, does not have any conventional IVs for education. The other dataset, which is for urban residents from across China, contains information on conventional IVs (parents' education). We find that, in the case of returns to schooling, the Lewbel method provides plausible estimates in datasets in which conventional IVs are not available and can be used to provide a robustness check on the findings from conventional IVs in datasets in which conventional IVs are available, but are weak. Our findings suggest that the method could prove useful in other contexts in which endogeneity is a problem and either conventional IVs are not available or those which are available may not satisfy the exclusion restriction.

1. Introduction

A large literature exists that estimates the returns to schooling using either ordinary least squares (OLS) or instrumental variables (IVs). The problem with relying on OLS estimates is that endogeneity and measurement problems potentially render simple OLS estimates inconsistent, making it is difficult to draw meaningful conclusions about the causal effect of education on earnings. The problem with obtaining IV estimates is that in many datasets conventional IVs are not available or, if potential candidates exist, they are often criticized for not satisfying the exclusion restriction. Our main contribution is that we illustrate how a novel identification strategy proposed by Lewbel (2012), which utilizes a heteroscedastic covariance restriction to construct an internal IV, can be used to obtain estimates of the returns to schooling. We illustrate how the Lewbel (2012) method can be used to obtain estimates for returns to schooling in two sets of circumstances. The first is with a dataset, in which there is no conventional IV for education at all. The second is with a dataset in which there are conventional IVs (parents' education), but they are potentially weak.

To illustrate our approach we use two datasets from urban China. One is a matched employer–employee dataset for Shanghai, in which there are no conventional IVs for education. The other is the China Household Finance Survey (CHFS), which is a nationally representative dataset for all of urban China, which contains information on parents' education. Given our focus on the returns to human capital and labor market outcomes, more generally we add to several recent papers that have been published in this journal on investment in human capital and the implications of such investment for wages, other labor market outcomes and economic growth. These include several related studies for China (Chen et al., 2014; Hertel and Zhai, 2006; Lei et al., 2013; Zhu et al., 2014) as well as for other developing Asian countries (see e.g. Jalil and Idrees, 2013; Pan, 2014; Qadri and Waheed, 2014; Sahoo and ten Raa, 2012).

There is a large literature that estimates the returns to schooling in China. Most of this literature, at least up until the mid-2000s, used OLS. More recently, a series of studies have used external IVs in an attempt to identify the causal effect of education on earnings in urban China (Chen and Hamori, 2009; Gao and Smyth, in press; Heckman and Li, 2004; Li and Luo, 2004; Mishra and Smyth, 2013). Common IVs used in the existing Chinese studies are parents' education (Heckman and Li, 2004; Mishra and Smyth, 2013); family background (Li and Luo, 2004); and spouse's education (Chen and Hamori, 2009; Gao and Smyth, in press; Mishra and Smyth, 2013).

Each of these IVs has problems in satisfying the exclusion restriction. Family background variables, including parents' education, potentially

☆ We thank the anonymous reviewers for their helpful suggestions on earlier versions of the paper.

* Corresponding author. Tel.: +61 3 99047179.
E-mail addresses: vinod.mishra@monash.edu (V. Mishra), russell.smyth@monash.edu (R. Smyth).
[1] Tel.: +61 3 99051560.

http://dx.doi.org/10.1016/j.econmod.2015.02.002
0264-9993/

have a direct effect on the individual's income level. For example, it can be argued that family background variables are correlated with family wealth, which then may have a direct influence on the income of their children. It may also be argued that family background variables are correlated with preference to find a job in a particular firm or industry, which then may have a direct influence on the individual's income.

People tend to partner with those who have similar individual endowments, such as education and other labor market characteristics. Thus, assortative mating has been used to justify spouse's education as an instrument. However, assortative mating also implies that spouse's education is likely to be highly correlated with family background variables and not really exogenous. If one uses spouse's education as the IV, there is also the disadvantage that the sample is restricted to those individuals who are married. These issues highlight the importance of alternative approaches to assessing the robustness of existing IV estimates of the returns to schooling in urban China and motivate the identification strategy employed in this study.

We stress the advantages of the Lewbel (2012) approach. The first advantage is that it can be used to provide IV estimates in cases in which other sources of identification, such as external IVs, are not available as is the case with the Shanghai dataset. The second advantage is that it can be combined with potentially weak IVs to provide more robustestimates in instances in which it is questionable whether the IVs that are available satisfy the exclusion restriction. This is the case with the CHFS, in which the best available IVs are parents' education. On the other side of the ledger, the Lewbel (2012) approach rests on certain assumptions, which we fully canvass in detail below, and because they are based on higher order moments, it is likely that they would not be as reliable as good conventional instruments.

2. Data

2.1. Shanghai dataset

The first sample is from a matched worker-firm data set from Minhang district in Shanghai, which was originally collected in 2007 by the Institute of Population and Labor Economics in the Chinese Academy of Social Sciences (CASS). The dataset was selected using Probability Proportion to Size sampling according to a list of all manufacturing firms in Minhang district with annual sales of at least 5 million RMB. The representativeness of the sample in terms of number of employees, sales revenue, profits and average wages are considered in Table 1. The firms in the sample are representative of firms in Minhang District and Shanghai as a whole.

The dataset contains information on 784 workers across 78 firms (on average 10.05 workers per firm). Once missing observations were removed, we had data on the variables of interest in this study for 628 workers across 72 firms. Table 2 provides descriptive statistics for the variables used in the study, based on the 628 workers from these 72 firms. The average hourly wage rate is 10.26 RMB. Among individual and human capital characteristics, 55.3% of respondents were male, 76.8% were married, 11.2% were members of the Chinese Communist Party, 58.3% held a non-agricultural *hukou* (household registration), the average years of experience was 16.1, and average years of schooling was 11.3. The firm characteristics are the proportion of female workers (average is 39%), proportion of migrant workers (average is 37%), whether there is a trade union presence in the firm (52.8% of firms have a trade union), firm size (average is 182 employees), exports/sales (0.13) and imports/sales (0.06) The dataset also contains information on the industry in which the firm is located (one of 22 manufacturing industries), as well as the ownership of the firm.

Table 1
Representativeness of the Shanghai sample.
Source: The data for Minhang District and Shanghai are from SBS (2008).

	Sample	Minhang District	Shanghai
Number of employee (person)	182.82	202.83	190.38
Sales revenue (10 thousand RMB)	8896.69	11,974.22	12,445.22
Profits (10 thousand RMB)	675.27	800.10	866.94
Average wage of employees (RMB/month)	2145.55	2383.42	2423.25

The Shanghai dataset has both advantages and disadvantages for our purposes. The main advantage is that it is matched data. Because we do have matched employer–employee data, we are able to provide estimates of returns to schooling controlling for individual as well as firm characteristics. The disadvantages are that the dataset is for a single city, it not right up-to-date and there are no conventional IVs with which to benchmark the IV estimates obtained using the Lewbel (2012) method.

2.2. China household finance survey

The second sample was from the CHFS. The CHFS was collected in 2011 by the Survey and Research Center for China Household Finance, which is located at the Southwestern University of Finance and Economics in China.[2] The CHFS employed a stratified three-stage probability proportion to size (PPS) random sample design. The first stage selected 80 counties (including county-level cities and districts) from 2,585 counties (primary sampling units, or PSUs) from all provinces and municipalities in Mainland China except Xinjiang, Tibet and Inner Mongolia. The second stage selected four neighborhood committees/villages from each of the selected PSUs at the first stage. The third stage selected 20–50 households (depending on the level of urbanization and economic development) from each of the neighborhood committees/villages chosen at the previous stage. Every stage of sampling was carried out with the PPS method and weighted by population size.

The 2011 CHFS collected information from households from rural and urban areas. In this study, we used a sample of 585 urban respondents who provided information pertaining to wages, education, parents' education and other labor market variables used in this study. Table 2 provides descriptives on the variables used in this study. Many of the control variables are the same as in the Shanghai dataset and, whenever possible, the same controls are used. The average hourly wage rate was 11.65 RMB, average years of schooling was 11.08 years, 71.8% of the sample were male, 75.9% were married, 12.1% were members of the Communist Party and 43.9% held a non-agricultural *hukou*. Among the two IVs, average father's schooling was 7.8 years and average mother's schooling was 6.02 years.

As with the Shanghai dataset, the CHFS dataset has its advantages and disadvantages. Advantages are that it is relatively recent, it is a representative dataset for all of urban China and it contains conventional IVs (parents' education). A disadvantage, compared to the Shanghai dataset, is that there is no data on employer characteristics.

3. Econometric methodology

We employ a Mincer earnings function in which gross hourly wage earnings including bonuses (measured in RMB) is regressed on years of schooling, post-school experience, post-school experience squared, and a series of control variables to capture individual characteristics (and firm-specific characteristics in the case of the Shanghai dataset). One should be careful in deciding whether to include job-specific and firm-specific characteristics in the earnings functions. Schultz (1988) made the point that wages and job-specific and firm-specific factors are likely to be jointly determined because some portion of educational returns is attributable to occupational choice. Our first specification employs years of schooling, post-school experience, post-school

[2] CHFS is publically available and can be accessed through registering at http://www.chfsdata.org.