Can Money ‘Buy’ Schooling Achievement?  
Evidence from 19 Chinese Cities

Guochang Zhao*

June 26, 2015

Abstract

This paper examines the causal effect of private tutoring on Chinese and mathematics test scores of primary school students in urban China. Because the unobserved determinants of schooling achievement often also influence private tutoring expenditure, the OLS estimate cannot provide a consistent estimate. This paper adopts a heteroskedasticity-based identification strategy proposed by Lewbel (2012) to handle this problem. The estimation results show that, on average, private tutoring expenditure has small but statistically significant effect on the mathematics test score of primary school students, but has no statistically significant effect on the Chinese test score. A 1000 yuan increase in private tutoring expenditure (i.e. 54% of a standard deviation) raises the primary school students’ mathematics test score by 1.07 percentage point (i.e. 15% of a standard deviation). The instrumental variable quantile regression combining with Lewbel IV suggests that private tutoring is more likely to improve student achievement at the bottom end of test score distribution. When moving upward to the top end, the effect becomes smaller and even negative despite not significant.

Key words: Private tutoring, shadow education, supplementary education, schooling achievement, heteroskedasticity

JEL classifications: I21 I24

*Research Institute of Economics and Management, Southwestern University of Finance and Economics. 55 Guanghua Cun Street, Chengdu, China, 610074. Email: guochangzhao@swufe.edu.cn
1 Introduction

Private tutoring (also known as shadow education or supplementary education), was particularly popular in East Asia. For example, in urban Japan over 90% of households bought private tutoring service for their children in 1997 (Bray, 2006); in the same year, 72.9% of primary school students, 56% of junior high school students and 32% of senior high school students in South Korea were receiving private tutoring (Kim and Lee, 2010). In recent years, its popularity has also grown quickly in other parts of Asia and in Africa, Europe, North America and Australia (Baker et al., 2001; Buchmann, 2002; Bray, 2006, 2007; Watson, 2008).

Given the increasing prevalence of private tutoring, a natural question is how effective it is. On the one hand, if private tutoring can effectively improve student achievement, the educational advantage of the families with better socio-economic status will be reinforced inter-generationally because the children from these families are more likely to participate in private tutoring (Bray and Kwok, 2003; Tansel and Bircan, 2006; Dang, 2007; Kang, 2011). This exactly runs counter to equality of opportunity. On the other hand, if private tutoring has little effect on schooling achievement, and therefore has little effect on final educational attainment, no matter how much households spend on private tutoring, it cannot be a policy issue in the sense of social equity. However, for this latter case, we will ask why the parents are willing to pay for private tutoring, and whether it is possible to improve a family’s utility, and thus the welfare of the whole society, with a different allocation of household spending. Therefore, for these two different scenarios, the policy implications of the effectiveness of private tutoring are quite different: for the first one, we should think how to relieve the education inequality due to the private tutoring; while for the second one, we need to inform parents the effectiveness of private tutoring and then they may switch the spending on private tutoring to something else, and improve the household utility in the end. In either case, estimating the causal effect of private tutoring on schooling achievement is the first step.

This paper aims to estimate the effect of private tutoring on student achievement in urban China. Private tutoring has been popular in urban China during the past decade or so, as in other areas of East Asia. Lei (2005) finds that in 2002-2003, about 37% of the third year senior high school students (Grade 12) participated in private tutoring.\footnote{About one third of the students in the data used by Lei (2005) are rural students.} Xue
and Ding (2008) find that the private tutoring participation rate among urban students was about 55.5% in 2004. We will see later that the RUMiC data revealed a similar private tutoring participation rate, and for participants, private tutoring expenditure, on average, accounts for three percent of household disposable income – or more than ten percent of family income per capita.

The major challenge in estimating the effect of private tutoring on school performance is the endogeneity problem. Because children’s innate abilities, drives, parents’ expectations and other factors are unobserved but can affect both private tutoring expenditure and children’s performances, the OLS estimates of the effect are not consistent. Earlier empirical studies (such as Stevenson and Baker (1992)) more or less ignore this problem. Recently Kang (2007a,b) attempts to use birth order as an instrumental variable to eliminate the bias; however, whether the birth order is a valid instrument is questionable.

In this paper, I firstly use a theoretical model to highlight the source of endogeneity associated with private tutoring - the omitted-variable bias. Then I adopt a heteroskedasticity-based method (Lewbel, 2012) to identify and estimate the effect of private tutoring on schooling achievement. Compared to the conventional IV estimation, this method does not need an excluded instrumental variable, however, it does need one or more variables, usually a subset of control variables, which are independent of the omitted variables or uncorrelated to the variance of the omitted variables; and it does require the error terms for private tutoring expenditure to be heteroskedastic with respect to these variables. In practice, this method can either be implemented by GMM estimation, which relies on a few high moment conditions, or a simpler estimator, which constructs instrumental variables based on the heteroskedasticity and then uses IV estimation. I adopt the latter one and call it Lewbel IV hereafter.

I am going to use the following variable to construct instrumental variables in Lewbel IV - city-grade\(^2\) average private tutoring expenditure. This variable should have nothing to do with student’s individual innate abilities, drive and other omitted variables. The error terms of private tutoring expenditure also have heteroskedasticity with respect to this variable. Therefore, it should satisfy the identification assumptions of Lewbel IV.

This paper uses the data from the 2008, 2009 and 2010 urban surveys of the Rural-Urban Migration in China (RUMiC) project — the RUMiC has a separate survey for

\(^2\)The ‘grade’ here refers to a group of classes rather than school performance rank.
rural-urban migrants, while the urban survey only includes the households with urban Hukou\(^3\); that is there is no migrants in my sample. Apart from a rich set of demographic measures, these data also record students’ Chinese and mathematics test scores in the final exam of last semester, along with household spending on formal education and private tutoring. About half of the students in the data participated in private tutoring, which can provide sufficient variation to obtain precise estimates.

The Lewbel IV estimation results suggest that on average private tutoring has significant but modest effect on the mathematics test scores of primary school students, but has no statistically effect on their Chinese test score. An 1000 yuan increase in private tutoring expenditure (i.e. 54% of a standard deviation) leads to an increase of 1.07 percentage point (i.e. about 15% of a standard deviation) in the test score. This is more than five times of the magnitude of the OLS estimate. These suggest that private tutoring expenditure is negatively related to the omitted variables, and the OLS estimate is downward biased.

The instrumental variables quantile regression results show that private tutoring has different effects over the distribution of test scores, and mainly the students in the bottom end of the test score distribution benefit from private tutoring. For example, when private tutoring expenditure increases by 1000 yuan, the mathematics test score increases by 1.9 percentage points (more than a quarter of a standard deviation) in the 0.10th quantile. When moving up to the upper quantiles, the effect reduces. For the Chinese test score, the results are a little mixed, but still suggest that private tutoring is more likely to improve test scores in the lower quantiles.

The rest of the paper is organized as follows. Section 2 summarizes the previous studies on private tutoring. Section 3 provides a theoretical model to help understand why and how private tutoring is endogenous. Section 4 introduces the empirical model and estimation strategies. Section 5 describes the data. The main empirical results and robustness checks are in the sixth and seventh sections, respectively. In section 8, I further discuss the possibility of reverse causality and the heterogeneity of causal effect of private tutoring on schooling achievement, and gauge the bias in OLS estimates using a panel sample and fixed effects model. Finally, section 9 concludes.

---

\(^3\)Hukou is a household registration system. There are only limited channels to get the urban Hukou, while in general rural-urban migrants have no access to these channels.
2 Literature review

The research on private tutoring escalated in recent years. This is because the earlier research on student achievement and education resources focused on Western countries, where the free compulsory schooling system was established long ago - for example in the early 1900s in the U.S. Over the past decades, private tutoring has become more and more popular all over the world (Bray, 2006; Watson, 2008) and the micro data sets have become available in East Asia, where private tutoring was firstly prevalent.

The education scholars firstly analyzed private tutoring. As summarized by Bray (2006), these educational studies try to understand private tutoring from different angles: quantitative patterns and variations, diversity in forms of supply, motives for seeking tutoring, effectiveness of tutoring in improving school performance and social stratification and government responses to private tutoring.

Economics research on private tutoring began more recently. It concentrates on the household demand for private tutoring – e.g. Bray and Kwok (2003) for Hong Kong, Tansel and Bircan (2006) for Turkey, Dang (2007) for Vietnam, Kim and Lee (2010) and Kang (2011) for South Korea. The common finding is that household income and parental education are positively related to private tutoring expenditure, while the number of siblings is negatively related to private tutoring expenditure. In contrast, the study on private tutoring in Canada by Davies (2004) finds that parents who hire or desire tutoring do not generally differ from other parents in their demographic or political ideology. However, parents who employ tutors are far more desirous of private schooling. Thus, he concludes that in Canada private tutoring represents an affordable alternative to private schools.

A number of studies have indicated that private tutoring could improve student achievement - for example, Stevenson and Baker (1992) for Japan, Mischo and Haag (2002) for Germany, Tansel and Bircan (2005) for Turkey, Ireson and Rushforth (2005) for the UK and so on. Some studies find no significant effect, e.g. Smyth (2008). Nevertheless, analogous to the studies about the effect of school resources on schooling achievement, estimating the causal effect of private tutoring on student achievement suffers from an endogeneity problem due to the omission of unobserved family and child characteristics. The empirical studies mentioned above do not pay much attention on this issue.

Kang (2007a,b) attempts to overcome the endogeneity problem using birth order as an
instrument. Using the standard IV estimation, Kang (2007a) finds that private tutoring expenditure has only a modest positive effect on student achievement in South Korea: a 10 percent increase in expenditure leads to a 0.56 percentage point improvement in test score. However, birth order may not be a valid instrument, because it may not satisfy the exclusion condition. As indicated, “a large body of literature theoretically and empirically documents that parents favor a certain-parity child (e.g., first-born or last-born) in education (Kang, 2007a).” This special preference may not only affect the child’s schooling expenditure, but also determines the time the parents are willing to spend on tutoring the child’s assignments or other tasks related to the child’s academic performance. Without controlling for these, the birth order is not a valid instrument. If the monetary expenditure on a child’s schooling is related to the parental time used on the child’s education or other parental help given to the child, this IV estimate is still inconsistent.

Kang (2007b) is aware of this problem and uses nonparametric bounds to overcome it. The advantage of the bounding method is that when there exists an endogenous variable, the instrumental variable only needs to be monotonically correlated to the error term but does not have to be uncorrelated to it. However, in Kang’s (2007b) situation, even this weak condition may not be satisfied. This is because some parents prefer the first-born while others prefer the last-born. Unless this order preference is the same for all families, the birth order cannot be monotonically related to the error term. In any case, using the bounding method, Kang (2007b) shows that the causal effect of private tutoring is at most modest: a 10 percent increase in private tutoring expenditure raises a student’s test score by 0.764 percentage point at the most.

As mentioned in the Introduction, this paper attempts to eliminate the bias relying on a heteroskedasticity-based method. Compared to the FE or IV estimations, this method neither requires the unobserved factors, which are correlated to private tutoring expenditure, to be time-invariant, nor needs an excluded instrument for identification. In contrast, the identification assumption of this method is based on high moments, which are more likely to be satisfied than the identification assumptions of FE and IV estimations in this study. I will show and explain this in the subsequent sections.
3 Theoretical model

This section presents a theoretical model adapted from Kim and Lee (2010). This model highlights how private tutoring (expenditure) is determined, and justifies the estimation strategy used in the empirical analysis. Consider a household $i$, which has a couple and only one child. This is a typical household in urban China because of the One-Child Policy. Assume that the household has a utility function of the form

$$U_i = s_i^{\alpha_1} (x_i - \underline{x}_i)^{\alpha_2},$$

where $s$ is the child’s schooling achievement, measured in school test scores; $x$ is a numeraire good, and $\underline{x}$ is the minimum consumption level on the numeraire good of the household; and $\alpha_1$ and $\alpha_2$ reflect the relative weights of child achievement and consumption in the household’s utility function, which are both assumed to be strictly positive.

Note that the utility function is based on an ‘altruistic assumption’, since the child’s schooling enters the household utility function just like the numeraire good but the parents do not benefit from the child’s education in this static model.

Furthermore, I also make these following assumptions. (i) A child’s schooling achievement is linearly determined by formal school education ($e^f$), private tutoring ($e^p$), family income ($y$) and other factors ($v$), e.g., parental education, the child’s innate abilities and drive, value the family places on education, teacher quality, the degree of diligence, accumulated knowledge, etc. (ii) All determinants of schooling achievements except private tutoring are exogenous, i.e., the couple can only choose different levels of private tutoring to influence their child’s achievement. (iii) The minimum consumption on the numeraire good is assumed to be a linear function of a vector of household and some neighborhood characteristics ($w$), which is also exogenous and may share some common factors with $v$.

The household’s optimization problem is

$$\max_{e_i^p, x_i} U_i = \left[ s_i \left( e_i^p, e_i^f, y_i, v_i', \Theta' \right) \right]^{\alpha_1} (x_i - \underline{x}_i)^{\alpha_2},$$

s.t. $x_i + p^f e_i^f + p^p e_i^p \leq y_i,$

$$s_i \left( e_i^p, e_i^f, y_i, v_i', \Theta' \right) = \beta^p e_i^p + \beta^f e_i^f + \gamma y_i + \theta' v_i,$$

$$\underline{x}_i = \phi' w_i,$$

$$\underline{x}_i > 0, x_i \geq \underline{x}_i, s_i > 0, e_i^f > 0, e_i^p > 0.$$
In the above model, inequality (2) is the budget constraint, in which the price of \( x \) is normalized to 1 since it is numeraire good, and \( p^f \) and \( p^p \) are the prices of formal school education and private tutoring, respectively. In the education market, households are assumed to be price takers, and prices are strictly positive. Equation (3) is the schooling achievement ‘production’ function and \( \Theta = (\beta^f, \beta^p, \gamma, \theta)' \). Equation (4) describes how the minimum consumption on the numeraire good is determined. Inequalities (5) present a set of restrictions, which set the boundaries of private tutoring within which the household could choose.

Given the above assumptions, the optimal demand for private tutoring, \( e^{p*}_i \), is

\[
e^{p*}_i = \begin{cases} 
0, & \text{if } \tilde{e}^{p}_i < 0; \\
\tilde{e}^{p}_i, & \text{if } 0 \leq \tilde{e}^{p}_i \leq \frac{y_i - p^f e^f_i - \phi' w_i}{p^p}; \\
\frac{y_i - p^f e^f_i - \phi' w_i}{p^p}, & \text{if } \tilde{e}^{p}_i > \frac{y_i - p^f e^f_i - \phi' w_i}{p^p};
\end{cases}
\]  

(6)

where \( \tilde{e}^{p}_i \) is the optimal demand for private tutoring without the restrictions presented in Inequalities (5), and its closed-form solution is

\[
\tilde{e}^{p}_i = \pi_1 y_i + \pi_2 e^f_i + \pi_3 w_i + \pi_4 v_i,
\]  

(7)

where \( \pi_1 = \frac{\alpha_1 \beta^p - \alpha_2 p^p}{(\alpha_1 + \alpha_2) \beta^p p^p}, \pi_2 = -\frac{\alpha_1 \beta^f + \alpha_2 p^p \beta^f}{(\alpha_1 + \alpha_2) \beta^f p^p}, \pi_3 = -\frac{\alpha_1}{(\alpha_1 + \alpha_2) p^p} \cdot \phi, \pi_4 = -\frac{\alpha_2}{(\alpha_1 + \alpha_2) \beta^f} \cdot \theta. \)

Ignoring the corner solutions, we can learn two lessons from Equations (3) and (7). First, if some factor included in \( v \) has an effect on schooling achievement, it must influence how much the household spends on private tutoring. This is the source of the endogeneity problem in empirical studies since not all of the factors included in \( v \) are observable. If we further assume the sign of the effect of private tutoring on schooling achievement, and assume that \( v \) and \( w \) do not share the same unobserved factors, we can know the direction of the bias in the OLS estimate for \( \beta^p \). However, in general, \( v \) and \( w \) often share some common factors.

For example, suppose private tutoring has a positive effect on schooling achievement, then although able children need less private tutoring to obtain some given achievement level, their parents often have higher expectations on their children’s achievement. Consequently, they might require a lower minimum level of the consumption of the numeraire good. Therefore, whether the OLS estimate for the causal effect is upward or downward biased is generally ambiguous.

---

4Although the Chinese central government announced in 2006 that the tuition and miscellaneous fees would be abolished from 2008 onward, the RUMiC data suggest that in urban China the charges by school only reduced about 14% from 2007 to 2008.
Second, the theoretical model shows that the schooling achievement production function, Equation (3), can be identified only if $w$ includes at least one unique variable that is not contained in $(y_i, e_i, v'_i)$. This is actually the exclusion condition in empirical studies. If this unique variable in $w$ is observable, it is the excluded instrument and the IV estimation is applicable in order to identify and estimate the causal effect of private tutoring on schooling achievement. If the unique variable is unobservable, we need some extra assumptions to identify the schooling achievement production function. I will discuss this in detail in the next section.

To sum up, the theoretical model points out the nature of the endogeneity problem, which is due to the omitted-variable bias, and also highlights the condition under which we might identify and estimate the causal effect of private tutoring on schooling achievement. However, this model also has the following limitations.

First, in the theoretical model the schooling achievement is measured using a comprehensive variable, but the fact is that students have to learn several different subjects at the same time, different students may attend different subjects’ private tutoring, and private tutoring in one subject might also influence the achievements in other subjects. In addition, children’s schooling achievement may not be the only issue parents care about. For example, fostering interests may be another purpose hiring private tutors. The theoretical model does not consider this situation. Second, the decision-makers in the theoretical model are parents and children’s reactions are not taken into account. Whether, and to what extent, the separation of participation decision-makers and participants influences the effect on schooling achievement is not modeled. Third, the theoretical model assumes that the effect of private tutoring is homogeneous across students and parents know how large the effect is when they make decisions on private tutoring participation and expenditure. The real situation is that the effect of private tutoring differs across children, and parents make decisions based on their expectations about the effect. This implies that in reality a negative effect of private tutoring on schooling achievement may exist for some students, but the theoretical model rules out this possibility. Because of these limitations, the predictions of the theoretical model are not necessarily the same as those from empirical work.

The theoretical model also has other minor limitations. For example, in the model schooling achievement is produced in a deterministic way, but in reality there are many
uncertainties in production of schooling achievement; the theoretical model does not explicitly distinguish the observed and unobserved variables; private tutoring is the only endogenous variable in the schooling achievement production function, while in reality there might be other endogenous variables such as formal education. Nevertheless, these limitations can be addressed directly in the empirical model.

4 Empirical model and estimation strategies

This section presents an empirical model, which parallels the schooling achievement production function (Equations (3)), and the demand function for private tutoring (Equation (7)) in the theoretical model. Rewriting these two equations, we get the following empirical model,

\begin{equation}
    s_i = \tau^p (p^p e^p_i) + \varphi'_1 X_i + \epsilon_{1i}, \quad \epsilon_{1i} = \vartheta_1 V_i + U_{1i};
\end{equation}

\begin{equation}
    (p^p e^p_i) = \varphi'_2 X_i + \epsilon_{2i}, \quad \epsilon_{2i} = \vartheta_2 V_i + U_{2i}.
\end{equation}

In this model, \(p^p e^p_i\) is the household spending on private tutoring. I substitute the private tutoring with the expenditure for private tutoring because the former is unmeasurable. Now the main objective is to consistently estimate \(\tau^p\), which equals \(\beta^p / p^p\). Next, \(X_i\) is a vector including family income \((y_i)\), formal education expenditure \((p^f e^f_i)\), and all observed factors included in \(v\) and \(w\). In the baseline model of this study, \(X_i\) is a vector of covariates including per capita family income, formal education expenditure, parental education, parental working status, child’s gender, number of siblings and fixed effects of school grade, residential city, survey year and school quality; in the augmented models, it also includes the city-grade average private tutoring expenditure.

For the error terms, \(V_i\) is the unobserved common determinants of schooling achievement and private tutoring expenditure such as child’s innate abilities, drive and so on. \(U_{2i}\) synthesizes the unobserved unique variables included in \(w_i\), which only influence private tutoring expenditure. Lastly, \(U_{1i}\) is the newly added term compared to the theoretical model. As mentioned in the last section, the theoretical model assumes that the schooling achievement is ‘produced’ in a deterministic way, but in the real world, there are many random shocks to examination results. For example, quiz setters always like to put unexpected questions in test papers, and even weather or missing breakfast in the examination day may have some influence on the examination results. All of these should
have little influence on the household’s decision on private tutoring, not to mention that
the decision is made at least several months before the examination — $U_{1i}$ captures this
type of unobserved factors. Also, $V_i$, $U_{1i}$, and $U_{2i}$ are assumed to be uncorrelated
with $X_i$ and are conditionally uncorrelated with each other, conditioning on $X_i$.

In the last section, the theoretical model indicated that private tutoring is endogenous
due to the omitted-variable bias. Thus, the OLS estimate for the causal effect of private
tutoring expenditure on schooling achievement, $\tau^p$, cannot be consistent. A typical way
to eliminate the bias is the instrumental variable approach. If at least one control variable
included in $X_i$ has no direct effect on schooling achievement (i.e. at least one element
in $\gamma_1$ is 0), then IV estimation can be applied. However, it is usually hard to find a
valid and strong instrumental variable. Alternatively, Lewbel (2012) suggests that if the
error term for private tutoring expenditure has heteroskedasticity, and a subset of the
control variables is uncorrelated to the variance of the common factor $V_i$ or independent
of the common factor $V_i$, some instrumental variable(s) can be constructed based on the
heteroskedasticity and an IV estimation (Lewbel IV) can be used to identify and estimate
the causal effect of private tutoring on schooling achievement.

Let us consider the two resolutions one at a time. First, the unobserved common
factor $V_i$ might include child’s innate ability, drive, parental expectation, accumulated
knowledge, degree of difficulty of curriculum, teacher quality, and so on. The IV es-

timation requires instrumental variables to have no correlation with these unobserved
factors but have strong effects on private tutoring expenditure. One potential candidate
of instrumental variable is city-grade average private tutoring expenditure.

The city-grade average private tutoring expenditure could be an instrument because
an individual’s behavior is more or less influenced by the collective behavior; however, the
collective behavior should be irrelevant to an individual’s unobservable characteristics.
The problem is that the purpose of an examination usually is not only to evaluate how
much knowledge students have learned but also to rank students’ performances. There-
fore, when the score is also used to rank students’ performance, it is more difficult for a
student to obtain higher test score if other students receive much private tutoring, unless
private tutoring has a negative effect or has no effect on schooling achievement. In other
words, the city-grade average private tutoring expenditure may have a direct effect on
individual’s test scores, and therefore may be invalid.
Second, the estimation and identification method based on heteroskedasticity (Lewbel, 2012) does not need excluded instruments, but it can consistently identify and estimate Equation (8) if the following assumptions are satisfied:

(i) $E(\varepsilon X) = 0$, where $\varepsilon = (\varepsilon_1, \varepsilon_2)$. That is, the control variables are uncorrelated with the error terms. This is the standard minimal regression assumption for the control variables.

(ii) $Cov(X, \varepsilon_2^2) \neq 0$, i.e., the error terms of Equation (9) must have heteroskedasticity with respect to the control variables.\(^5\)

(iii) $Cov(Z, \varepsilon_1 \varepsilon_2) = 0$ for an observed $Z$, where $Z$ can be a subset of $X$ or exactly equal to $X$. Based on the assumptions of the model, a sufficient condition to make this assumption hold is that the variance of the common factor is uncorrelated to $Z$, i.e., $Cov(Z, V^2) = 0$. A stronger but easily interpreted condition is that the common factor is independent of $Z$.

These assumptions imply the following moments

$$E[X\varepsilon_1] = 0, \ E[X\varepsilon_2] = 0, \ Cov(Z, \varepsilon_1 \varepsilon_2) = 0,$$

along with the heteroskedasticity of $\varepsilon_2$. Generally, the moment conditions (10) provide the basis of the GMM or GEM estimation, but in the case of omitted-variable bias (as in this paper) there is another simpler estimator: first linearly regress $p\rho p$ on $X$, then obtain the residuals $\hat{\varepsilon}_2$, and lastly estimate Equation (8) by IV estimation using $X$ and $(Z - \bar{Z})\hat{\varepsilon}_2$ as instruments, where $\bar{Z}$ is the mean of $Z$. Note that $Z$ is not necessarily a subset of $X$, and it can be just an excluded variable such as the classical instrumental variable, although it is often taken from the control variables.

Lewbel (2012) illustrates that if the endogeneity does not come from reverse causality, this method can easily extend to the cases when excluded instruments are available and when there is more than one endogenous variables. He also points out that, as the identification and estimation method is based on higher moments, it may provide less reliable estimates than identification based on the standard exclusion condition, but it may be useful in applications where traditional instruments are weak or non-existing. Thus, during the past decade, quite a few papers have used this method as the main or extra strategy to overcome the endogeneity problem (see, e.g. Giambona and Schwienbacher, 5).

---

\(^5\)If there exists reverse causality, Lewbel’s method also needs $Cov(X, \varepsilon_1^2) \neq 0$. In section 8, I will discuss the possibility of reverse causality and the influences on estimation results.
2007; Rashad and Markowitz, 2007; Sabia, 2007a,b,c; Emran and Hou, 2008; Huang et al., 2009), none of which, however, is about private tutoring.

In my case, whether the error term $\varepsilon_{2i}$ has heteroskedasticity or not is an empirical issue, which will be checked later. The discussion about the validity of the potential instrumental variable suggests that the city-grade average private tutoring expenditure should be independent of the unobserved characteristics such as the child’s innate abilities and drive, degree of difficulty of curriculum, accumulated knowledge, teacher quality and so on. This means that this variable can serve as $Z$ and thus this identification method based on heteroskedasticity is the main estimation strategy in the empirical analysis.

Before turning to the empirical results, I should note three points about the model specification. First, Equation (7) in the theoretical model is for the latent variable, while in Equation (9) the dependent variable is the actual private tutoring expenditure, which is censored at 0 in reality. Thus Equation (9) should be in a nonlinear functional form, but I adopt a linear functional form here for simplicity.\(^6\)

Second, private tutoring expenditure is measured in level. It is often taken a (natural) log if a variable, especially the dependent variable, is a positive amount of money for several advantages (Wooldridge, 2009, pp: 183-185). When the variable, for example $y$, can equal 0, log($1+y$) is used sometimes, and we can still interpret the estimates as if the variable were log($y$) when the data on $y$ are not dominated by zeros (Wooldridge, 2009, pp: 183-185). However, in this study about half students spend zero money on private tutoring. The estimation based on the log of private tutoring expenditure plus one when there exist many zero-value observations depends on unit of measurement, which is chosen arbitrarily. In addition, the main estimation strategy relies on the heteroskedasticity of the error terms of private tutoring expenditure. When the private tutoring expenditure takes log, the heteroskedasticity reduces significantly. Thus, this study uses the level rather than the log of the variable.\(^7\)

Third, the outliers will be identified based on some statistic and excluded in the estimation in order to avoid the influence of extreme values. Details of this are presented in the next section.

---

\(^6\)A Monte Carlo simulation shows that this simplification does not affect the estimation of Equation (8). The simulation program and results are available upon request.

\(^7\)Quite a few previous studies also use the level of expenditure or income rather than the log of these in modeling schooling achievement, see for example Kang (2007a), Houtenville and Conway (2008) and Dahl and Lochner (2012).
5 Data

The data used in this paper are from the Rural-Urban Migration in China (RUMiC) project conducted by The Australian National University. The RUMiC project has three components: rural survey, urban survey and rural-urban migration survey. Here, I use the urban survey, which only contains the urban households but no rural-urban migrants, i.e., all of households have urban Hukou. The advantage of this dataset is that it not only has detailed demographic information but also includes information about household expenditure on children’s education and children’s schooling achievement, both of which are key variables in this study.

The RUMiC is a four-year longitudinal survey from 2008, and the first three waves’ data for urban China were available at the time of writing. The urban survey of the RUMiC project was designed to contain 5000 households in 19 cities. In the second wave (2009), 285 original households were lost due to sample attrition and the same number of new households were supplemented. In the third wave (2010), the cities of Anyang and Jiande exited from the survey which resulted in the loss of 200 households, while 870 households were lost in the remaining 17 cities due to sample attrition, and another 870 households were supplemented. Consequently, the survey includes 4,800 households in urban China in the third wave.

Private tutoring expenditure is the key explanatory variable in this study, and just as the name implies it is paid to private tuition providers. The private tutoring expenditure usually refers to the spending on academic learning, but in the RUMiC data the training fees in arts or sports (if there are) are also added into private tutoring expenditure.

A shortcoming of the private tutoring expenditure in the RUMiC data is that it has no detailed information about the amount of private tutoring expenditure on different curriculum subjects. To further understand this issue, I conducted a small survey of private tutoring participation and expenditures on different subjects. The survey covers two primary schools in urban areas of Shaanxi Province. Appendix A presents more detailed information on this survey.

Table A.1 in Appendix A presents the summary statistics of this small survey. It shows that about 30% of students participate in the private tutoring in Chinese and

---

8These 19 cities are: Shanghai, Nanjing, Wuxi, Hangzhou, Ningbo, Jiande, Hefei, Bengbu, Zhengzhou, Luoyang, Anyang, Wuhan, Guangzhou, Shenzhen, Dongguan, Chongqing, Chengdu, Mianyang and Leshan.
mathematics. English is the most popular subject, with more than twice of the participation rate of Chinese and mathematics. For the participants, the average expenditures on Chinese and mathematics are also basically the same, while the average expenditures on English are much higher. If this is also true in the RUMiC data, is is implied that that while we focus on Chinese or mathematics test scores, in reality the private tutoring expenditure defined in the RUMiC data is larger than the actual respective private tutoring expenditures on these two subjects. This can be viewed as a systematic measurement error, and it leads to some problem in interpreting the results. I will discuss this in detail in the empirical results.

Chinese households need or may need to pay another two types of expenditure to schools, which are also available in the RUMiC data. One is tuition and related fees. These fees include tuition and miscellaneous fees, special tutorial fees, catering and accommodation fees if the students eat and live at school and others such as uniform fees. The other one is school choice fee. In urban areas, students are usually assigned to the nearest schools at levels of primary and lower-secondary education, while the enrolment to senior high school is based on the entrance examination. If a student’s Hukou is not in the school district where his/her school is located, or his/her test score is lower than the cutoff score in the entrance examination, he/she needs to pay a fee to be admitted to his/her school. Overall, these two types of expenditure are paid to formal schools, so I add them up and name it ‘formal education expenditure’.9

Student achievement is the dependent variable in this study. In the first wave, only the parents’ subjective assessment on children’s school performance is reported. Specifically, the question about the subjective assessment is “How is his/her performance in the class currently or before dropping out? (1) Excellent, (2) good, (3) average, (4) bad or (5) very bad.” That is the subjective assessment is an ordinal variable. From the second wave onward the parents are also required to report children’s actual and full Chinese language and mathematics test scores in the final examination of the last semester.10 Compared to the subjective assessments, the test score is a more precise and accurate presentation of school performance for two reasons. First, in China, it is very common for schools to send or post the transcripts to parents after the final examination. Also,

---

9 I also try to exclude the catering and accommodation fees from the formal education expenditure, and find that including them or not nearly has no impact on the estimation results.

10 The full score refers to the maximum test score students may get if they do not make any mistake in the examination.
the school usually requires students to return the transcripts with parents’ signatures. Second, less than two percent of the parents report that their children’s performance is bad or very bad in the RUMiC data. This suggests that parents may to some extent overestimate their children’s performance. The strength of subjective assessment is that it is a comprehensive evaluation, which might contain all sorts of achievements that children can have in the class besides Chinese and mathematics test scores, such as performances in English, even music, painting or sports lessons. This is the comparative advantage of the subjective assessment, and in this sense it may be better and more accurate than the test score, especially there is only a ‘total’ private tutoring expenditure in the data. However, the subjective assessment is an ordinal variable, while whether the estimation strategy - Lewbel IV - is applicable to the ordinal dependent variable is not clear yet. Thus, on the one hand, the main empirical analysis in this study uses the test score to measure student’s achievement; on the other hand, I classify the subjective assessment into a binary variable (being excellent or not) or simply view it as a continuous variable, and replicate the estimations as robustness checks — the results have a similar pattern with those of test scores.

Since the full test score varies over time, across cities and schools, to make the test scores comparable, I normalize the test score as \( \frac{\text{actual score}}{\text{full score}} \times 100 \). So one unit increase is one percentage point increase. The normalized test score measures approximately how many fractions of knowledge the students have mastered compared to the maximum level of tested knowledge.

This paper focuses on the primary school students (Grade 1-6) in urban China. To examine the effect of private tutoring on primary school student achievement, I first restrict the sample to children in primary schools with valid information on their Chinese or mathematics test scores and private tutoring expenditure. This results in 1,398 observations. Then I further exclude those with missing data in parental education, parental working status and school quality. The sample size reduces to 1,135.

Next, using the block adaptive computationally efficient outlier nominator algorithm (BACON) proposed by Billor et al. (2000), 15 outliers are identified as shown in Appendix Figure B.1. They are excluded in the following empirical analysis. Consequently, there are 1,120 observations (826 students) in the remaining sample (full sample). In this sample, 294 students were observed twice between 2009 and 2010 data (panel sample).
Table 1: Main sample characteristics of private tutoring participants and non-participants

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Non-participants</th>
<th>Participants</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (1)</td>
<td>S.D (2)</td>
<td>N (3)</td>
</tr>
<tr>
<td>Chinese test score</td>
<td>90.95</td>
<td>7.77</td>
<td>504</td>
</tr>
<tr>
<td>Mathematics test score</td>
<td>91.86</td>
<td>7.73</td>
<td>508</td>
</tr>
<tr>
<td>Formal education expenditure (yuan)</td>
<td>1515</td>
<td>2991</td>
<td>512</td>
</tr>
<tr>
<td>Private tutoring expenditure (yuan)</td>
<td>0</td>
<td>0</td>
<td>512</td>
</tr>
<tr>
<td>Father’s years of schooling</td>
<td>11.83</td>
<td>3.04</td>
<td>512</td>
</tr>
<tr>
<td>Mother’s years of schooling</td>
<td>11.07</td>
<td>2.86</td>
<td>512</td>
</tr>
<tr>
<td>Family annual income per capita (yuan)</td>
<td>18234</td>
<td>13813</td>
<td>512</td>
</tr>
<tr>
<td>Father: unemployed, retired or housekeeper</td>
<td>0.09</td>
<td>0.05</td>
<td>608</td>
</tr>
<tr>
<td>Father: paid employment</td>
<td>0.78</td>
<td>0.84</td>
<td>608</td>
</tr>
<tr>
<td>Father: self-employment</td>
<td>0.13</td>
<td>0.12</td>
<td>608</td>
</tr>
<tr>
<td>Mother: unemployed, retired or housekeeper</td>
<td>0.24</td>
<td>0.19</td>
<td>608</td>
</tr>
<tr>
<td>Mother: paid employment</td>
<td>0.66</td>
<td>0.75</td>
<td>608</td>
</tr>
<tr>
<td>Mother: self-employment</td>
<td>0.11</td>
<td>0.07</td>
<td>608</td>
</tr>
<tr>
<td>Number of siblings</td>
<td>0.25</td>
<td>0.49</td>
<td>512</td>
</tr>
</tbody>
</table>

Note: The income and expenditure are deflated to the price level of 2009 using provincial-specific urban CPIs. Data source: The Rural-Urban Migration in China (RUMiC) survey.

Given that nearly half of students are lost in the panel sample, the full sample rather than the panel sample is used in the main empirical analysis, while after the main empirical analysis I will compare the results of the panel sample with those of the full sample and further discuss their implication.

Table 1 describes characteristics of the full sample based on private tutoring participation.\(^ {11}\) It shows that the private tutoring participants have higher test scores, also spend more on formal education, have more educated parents and fewer siblings; their families are richer and their parents, especially mothers, are less likely to be unemployed, retired or a housekeeper. Overall, the table illustrates that the private tutoring participants have higher test scores and a better socio-economic background than non-participants.

Figure 1 plots the unconditional relationships between test scores and private tutoring expenditure using the local mean smoothing method. It illustrates that the mathematics test score is positively associated with private tutoring expenditure, but the Chinese test score does not exhibit a strong relation to private tutoring expenditure.

Columns (1) to (3) of Appendix Table C.1 display summary statistics for the full

\(^ {11}\)In the table, the family income is defined as total income minus the taxes and transferred expenses. This information is unavailable for approximately 13% of the observations. In this case, the family income is imputed by the students’ parental income and other parental demographic characteristics controlling for the residential city.
Figure 1: Unconditional relationship between private tutoring expenditure and test score

Note: The private tutoring is measured at the 2009 price. The unconditional relationships are estimated using the local mean smoothing method. Data source: The Rural-Urban Migration in China (RUMiC) survey.

The city-grade average private tutoring expenditure refers to the average of the private tutoring expenditure of other students in the same city and same or adjacent school grade. The adjacent school grade for a particular student mean the grade above and below the grade which the student was in. For example, the adjacent grades for a Grade-3 student are Grades 2 and 4. For students in Grades 1 and 6, there is only one adjacent grade for them. They are Grades 2 and 5, respectively. Including the students in adjacent grades to calculate the city-grade average private tutoring expenditure is because in a few city-grade cells, there are only fewer than 10 observations. In addition, although the sample in empirical analysis requires all of the variables nonmissing, all of the observations which have valid private tutoring information and are not identified as outliers are used to calculate the city-grade average private tutoring expenditure. As mentioned earlier, this variable is used to construct the instrumental variable in the Lewbel IV estimation.

The variable ‘school quality’ is a subjective assessment reported by parents. The
question in the survey is “what is the education quality of the school he/she is attending currently or before dropping out? (1) The best in the city, (2) better than average in the city, (3) average or (4) worse than average in the city.” Thus, school quality has four ranks. In the sample only 0.34 percent of parents believe that their children are at worse than the average schools. Therefore I merge the four categories into three: the best, better than average and average or below. The asymmetrical distribution of these subjective assessments illustrates that parents are optimistic to some extent about their children’s school quality. This optimism is not rare in the literature about subjective assessments (e.g. Groot, 2000; Lokshin and Ravallion, 2008). However, without more objective measures, this one is the optimal we have to control for school quality.

Finally, note that there are not many students in Grade 1 compared to the other school grades. The 2010 China Census data indicates that the population size of the birth cohorts who should have begun primary education in 2009 and 2010 is about 10 percent less than the population size of those who should have begun primary education during the previous 5 years (NBS of China, 2012). This may be why the students in Grade 1 are fewer than those in other grades.

6 Empirical results

In this section, the effects of private tutoring on Chinese and mathematics test scores are estimated with OLS and Lewbel IV. The OLS is the benchmark. Although the OLS estimate is inconsistent because of the omitted-variable bias, it could tell us what the relationship between schooling achievement and private tutoring looks like, while the Lewbel IV estimations can eliminate this bias.

6.1 OLS estimation results

Columns (1) and (3) in Table 2 present the OLS results for the schooling achievement equation. They show that the private tutoring expenditure is positively associated with test score, and while the association is significant for mathematics, this is not the case for Chinese. These results are consistent with the unconditional relationships between private tutoring expenditure and test scores presented in Figure 1. In terms of the magnitude, the association is quite modest. Columns (1) and (3) indicate that every additional 1000
yuan spent on private tutoring is associated with a 0.12 and 0.20 percentage point change in Chinese and mathematics test scores, respectively. These changes correspond to a less than three percent of standard deviation of the test scores.

Note that the test scores come from different tests. This fact may lead to some measurement error in student achievement though the curriculum is the same nationwide. Since this measurement error happens to the dependent variable, it reduces the estimation efficiency but does not cause attenuation bias. In order to take into account of this random effect, I also try to adjust the standard error by clustering at district level - the examination is usually the same within district. As a result, the standard error changes a bit but the significance level does not change, and this is the same for the Lewbel IV estimates. Given that the city fixed effect is controlled for, these results may imply that the difficulty of examinations is basically the same within a city.

The OLS results in Table 2 also show that formal education expenditure has neither an economically substantial nor a statistically significant relation with test scores - this might be because how much a household pays for formal education is mainly decided by the school. Girls perform better in Chinese, but in mathematics there is no significant difference between boys and girls, which is consistent with recent studies such as Guiso et al. (2008). School quality matters to student achievement, but the difference between the categories ‘the best school’ and ‘above average school’ is not significant statistically. Test scores are negatively and statistically associated with school grades. This is expected, since in China the difficulty level of the curriculum increases with grades and the function of differentiating students’ learning outcomes is also emphasized with grades.

6.2  Lewbel IV estimation results

This subsection presents the Lewbel IV estimation results. As described in the section of methodology, I use the city-grade average private tutoring expenditure to construct the instrumental variable.

6.2.1  The first stage results

Lewbel IV estimation is based on the constructed instrumental variable, which relies on the heteroskedasticity of the error term with respect to some exogenous variable (Z) in the first stage. Here the variable Z is the city-grade average private tutoring expenditure.
Table 2: OLS and Lewbel IV estimation results of the schooling achievement equation

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Chinese</th>
<th>Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS (1)</td>
<td>Lewbel IV (2)</td>
</tr>
<tr>
<td>Private tutoring expenditure (1000 yuan)</td>
<td>0.115 (0.105)</td>
<td>0.562 (0.443)</td>
</tr>
<tr>
<td>Formal education expenditure (1000 yuan)</td>
<td>0.005 (0.052)</td>
<td>-0.003 (0.056)</td>
</tr>
<tr>
<td>City-grade average private tutoring expenditure (1000 yuan)</td>
<td>-0.734 (0.493)</td>
<td>-0.682 (0.487)</td>
</tr>
<tr>
<td>Per capita family annual income (1000 yuan)</td>
<td>0.038** (0.019)</td>
<td>0.033* (0.019)</td>
</tr>
<tr>
<td>Father’s years of schooling</td>
<td>0.126 (0.096)</td>
<td>0.115 (0.096)</td>
</tr>
<tr>
<td>Mother’s years of schooling</td>
<td>0.165* (0.094)</td>
<td>0.131 (0.098)</td>
</tr>
<tr>
<td>Father: paid employment</td>
<td>0.518 (0.820)</td>
<td>0.316 (0.830)</td>
</tr>
<tr>
<td>Father: self-employment</td>
<td>-0.097 (1.035)</td>
<td>-0.326 (1.046)</td>
</tr>
<tr>
<td>Mother: paid employment</td>
<td>-0.024 (0.493)</td>
<td>0.060 (0.497)</td>
</tr>
<tr>
<td>Mother: self-employment</td>
<td>-1.462 (0.983)</td>
<td>-1.448 (0.969)</td>
</tr>
<tr>
<td>Female student</td>
<td>1.744*** (0.416)</td>
<td>1.615*** (0.436)</td>
</tr>
<tr>
<td>Number of siblings</td>
<td>-0.910 (0.739)</td>
<td>-0.755 (0.738)</td>
</tr>
<tr>
<td>School quality: above average</td>
<td>-1.095** (0.551)</td>
<td>-1.047* (0.554)</td>
</tr>
<tr>
<td>School quality: average or below</td>
<td>-2.426*** (0.633)</td>
<td>-2.362*** (0.631)</td>
</tr>
<tr>
<td>Grade 2</td>
<td>-0.905 (0.643)</td>
<td>-0.969 (0.642)</td>
</tr>
<tr>
<td>Grade 3</td>
<td>-3.429*** (0.653)</td>
<td>-3.587*** (0.676)</td>
</tr>
<tr>
<td>Grade 4</td>
<td>-2.926*** (0.658)</td>
<td>-3.248*** (0.726)</td>
</tr>
<tr>
<td>Grade 5</td>
<td>-3.709*** (0.715)</td>
<td>-4.071*** (0.805)</td>
</tr>
<tr>
<td>Grade 6</td>
<td>-5.604*** (0.816)</td>
<td>-5.880*** (0.870)</td>
</tr>
<tr>
<td>Constant</td>
<td>88.374*** (1.806)</td>
<td>88.660*** (1.822)</td>
</tr>
</tbody>
</table>

Cragg-Donald Wald F statistic — 75.66 — — 78.31
Kleibergen-Paap rk Wald F statistic — 19.36 — — 20.16
Observations 1109 1109 1111 1111

Note: Robust standard errors are in parentheses, and are clustered at the person level. *** p<0.01, ** p<0.05, * p<0.10. All estimations control for city and year fixed effects. Data source: The Rural-Urban Migration in China (RUMiC) survey.
Columns (1) and (3) of Appendix Table C.2 present the regression results of private tutoring expenditure on all control variables for Chinese and mathematics samples, respectively. The coefficient of the city-grade average private tutoring is negative, but the magnitude is quite small and insignificant at all. Therefore, even if the city-grade average private tutoring expenditure satisfies the exclusion condition, it is only a very weak instrument. The Koenker heteroskedasticity test statistic\textsuperscript{12} suggests that the error term is strongly heteroskedastic with respect to city-grade average private tutoring expenditure. Thus, we can expect that the constructed instrumental variable should be highly correlated with the endogenous variable - private tutoring expenditure.

Using the residuals ($\hat{\varepsilon}_2$) in Columns (1) and (3), an instrumental variable is constructed as $(Z - \bar{Z})\hat{\varepsilon}_2$ (call it CIV, hereafter), where $Z$ is the city-grade average private tutoring expenditure and $\bar{Z}$ is its sample mean. Columns (2) and (4) of Appendix Table C.2 present the first stage results using CIV as the instrumental variable. The Cragg-Donald Wald F-statistic (hereafter, CD F-statistic) suggests that the strength of this instrument is quite strong; the Kleibergen-Paap rk Wald F-statistic (hereafter, KP F-statistic), which drops the i.i.d. assumption imposed in the CD F-statistic and is based on clustered variance-covariance matrix in this study, drops a lot, but according to the ‘rule of thumb’, the CIV is still a strong instrumental variable\textsuperscript{13}.

In addition, the other important characteristic of the first stage results, presented in Appendix Tables C.2, is that formal education expenditure is positively related to private tutoring expenditure, especially when CIV is controlled for, although the relation is weak in terms of magnitude. Indeed, this result is in contrast to the theoretical model. The theoretical model suggests that when formal education expenditure is higher, the private tutoring expenditure should be lower (see Equation (7)). The behind intuition is that formal education and private tutoring are substitutes.

One potential reason for this difference is that formal education (expenditure) is assumed to be exogenous in the theoretical model, while in reality there may exist some room for discretionary formal education expenditure for households, such as special tuto-

\textsuperscript{12}This test drops the assumption of normal distributed error terms imposed in the Breusch-Pagan test, and thus is the robust analog of the latter. The null hypothesis of the Koenker heteroskedasticity test is that the error term is homoskedastic.

\textsuperscript{13}The critical values for weak IV tests in the presence of non-i.i.d. errors have not been compiled yet. However, Baum et al. (2007) advise that “users either apply with caution the critical values compiled by Stock and Yogo (2005) for the i.i.d case or refer to the older ‘rule of thumb’ of Staiger and Stock (1997), which says that the F-statistic should be at least 10 for weak identification not to be considered a problem.”
rials and school choice fees. Therefore, how much the households spend on their children’s private tutoring and formal education may be influenced by some common factors such as the value the family places on education. Hence, these two expenditures are positively related to each other. In the section on robustness check, formal education expenditure is allowed to be endogenous, and we can find that this change has no substantial impact on the estimated effect of private tutoring on student achievement.

6.2.2 The second stage results

Columns (2) and (4) of Table 2 presents the second stage (Lewbel IV estimation) results. The results indicate that private tutoring expenditure has no statistically significant effect on the Chinese test score, while the effect on the mathematics test score is positive and statistically significant: an extra 1000 yuan spent on private tutoring increases the mathematics test score by 1.07 percentage point. Given that the standard deviation of mathematics test score is 7.35, this 1.07 percentage point is a modest effect. However, compared to the OLS estimate, where an extra 1000 yuan in private tutoring expenditure can only raise the mathematics test score by about 0.2 percentage point, IV estimate is more than five times of the OLS estimate. Kang (2007a,b) also finds similar results that IV or IV bounding estimates are higher than the OLS estimates in South Korea - but the score he uses is the average of test scores of mathematics, Korean language and English.

Now, two questions arise from these results. One is why private tutoring has different effects on Chinese and mathematics test scores; the other one is why private tutoring becomes increasingly prevalent given the nil or modest effect of private tutoring on test scores.

An explanation for the first question may be that few students participate in Chinese private tutoring, and thus the (total) private tutoring expenditure used here is not relevant to the Chinese learning outcomes. However, the survey I conducted in Shaanxi Province shows that the participation rates and private tutoring expenditure in Chinese and mathematics are basically the same (Table A.1 in Appendix A), although the RU-MiC data have no the detailed information on expenditure for different subjects. Thus, this argument may not hold. Another explanation may be that for Chinese learning, the test score is not the only, or main, outcome which parents care about. If so, the content of private tutoring in Chinese may not be directly relevant to the test score. However,
because we have no information about the content of private tutoring, this is only a conjecture.

For the second question, the above conjecture about the purpose for parents hiring private tutoring for their children may also give some explanation. When the main purpose is not to improve a child’s test score, but something else such as the child’s interests, the nil or modest effect on the test score is a natural result. Again, because of the limitation of the data, we cannot further examine this possibility.

Another possible explanation for the second question is that whether or not to attend private tutoring is a decision made by parents rather than by the children themselves. The RUMiC data have no information on this either, but the small survey I conducted indicates that more than 40% of students are a little, or very, reluctant to attend private tutoring and less than 20% of students are very willing to attend private tutoring. If children are not willing to do so, and therefore are not cooperative with the private tutors, the effect should be discounted substantially.

The last explanation for the second question is about the definition of private tutoring expenditure in the RUMiC data. As mentioned above, private tutoring expenditure in the data is a total expenditure and we do not know the specific amount spent on Chinese, mathematics or indeed anything else. Therefore, suppose for a student that private tutoring expenditure increases by 1000 yuan, the money used on a particular subject should be a part of this 1000 yuan. Take mathematics as an example, according to the small survey I conducted in Shaanxi Province, the expenditure for mathematics private tutoring is about a quarter of the total private tutoring expenditure on average. Thus, a naive calculation implies that if all the 1000 yuan is spent on mathematics, the child’s mathematics test score increases by 4.3 percentage points (i.e. about 58% of a standard deviation) – this is a relative large number. However, the reality should be much more complex than this naive scenario: what, and how many subjects have private tutoring, as well as the expenditure for each subject, differ across students; and private tutoring for one subject may have a positive or negative effect on other subjects. Therefore, based on this, we can only speculate that the effect of an additional 1000 yuan spent on mathematics private tutoring should be larger than the estimate obtained here, but the extent to which the actual effect is greater is unknown.
Table 3: Robustness checks (1)

<table>
<thead>
<tr>
<th></th>
<th>Chinese OLS</th>
<th>Chinese Lewbel IV</th>
<th>Mathematics OLS</th>
<th>Mathematics Lewbel IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Panel 1: Control for birth weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private tutoring expenditure (1000 yuan)</td>
<td>0.056</td>
<td>0.170</td>
<td>0.150</td>
<td>0.764**</td>
</tr>
<tr>
<td></td>
<td>(0.112)</td>
<td>(0.440)</td>
<td>(0.114)</td>
<td>(0.365)</td>
</tr>
<tr>
<td>Kleibergen-Paap rk Wald F-statistic</td>
<td>—</td>
<td>18.03</td>
<td>—</td>
<td>19.25</td>
</tr>
<tr>
<td>Observations</td>
<td>1,027</td>
<td>1,027</td>
<td>1,028</td>
<td>1,028</td>
</tr>
<tr>
<td>Panel 2: Treat formal education expenditure as another endogenous variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private tutoring expenditure (1000 yuan)</td>
<td>0.116</td>
<td>0.250</td>
<td>0.200*</td>
<td>0.670***</td>
</tr>
<tr>
<td></td>
<td>(0.106)</td>
<td>(0.274)</td>
<td>(0.102)</td>
<td>(0.226)</td>
</tr>
<tr>
<td>Formal education expenditure (1000 yuan)</td>
<td>0.006</td>
<td>0.041</td>
<td>0.036</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.056)</td>
<td>(0.044)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>Kleibergen-Paap rk Wald F-statistic</td>
<td>—</td>
<td>16.57</td>
<td>—</td>
<td>16.57</td>
</tr>
<tr>
<td>Observations</td>
<td>1,109</td>
<td>1,109</td>
<td>1,111</td>
<td>1,111</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses, and are clustered at the person level. *** p<0.01, ** p<0.05, * p<0.10. All specifications control for per capita family annual income, parental education, parental working status, gender, number of siblings, fixed effects of school grade, school quality, residential city and city-grade average private tutoring expenditure. Birth weight and city-grade average formal education expenditure are also included as additional control variables, respectively, in panels 1 and 2. Data source: The Rural-Urban Migration in China (RUMiC) survey.

7 Robustness checks

I conduct three robustness checks in this section.

First, I add an additional control variable - birth weight, to the schooling achievement equation. The literature has found that birth weight captures the initial health and some other endowments, and it might have a long-term effect on educational attainment (e.g. Black et al., 2007). Thus, this variable also influences private tutoring expenditure. If this is the main sources of endogeneity, controlling for them might make the OLS estimates close to the IV estimates. Panel 1 of Table 3 presents the results with this additional control variable. It indicates that the estimation results are basically the same as the main results.

Second, I consider the case that formal education expenditure may be also endogenous. Recall that the theoretical model predicts that formal education (expenditure) has a negative effect on private tutoring (expenditure) when the former is exogenous, while the first stage results presented in Tables C.2 show that formal education expenditure may

14Refer to my theoretical model in section 3 - all the (expected) factors affecting schooling achievement are also determinants of private tutoring expenditure.
have a positive effect on private tutoring expenditure. As discussed for the first stage results, this might be because that formal education expenditure is partially determined by the household, for example, school-choice fee. If so, formal education expenditure is also related to the omitted variables.

Panel 2 of Table 3 presents the estimation results under the assumption that formal education expenditure and private tutoring expenditure are both endogenous. I use the city-grade average formal education expenditure to construct an instrumental variable for an individual’s formal education expenditure, as was done for private tutoring expenditure. Compared to the case when only private tutoring expenditure is endogenous, first, the KP F-statistic reduces a bit. This is because the formal education expenditure is less heteroskedastic, while the Lewbel IV method constructs the instrumental variables based on heteroskedasticity. Second, the Lewbel IV estimate reduces, but the pattern remains the same. That is, assuming formal education expenditure to be endogenous does not have much impact on the estimation results.

Third, I use parental subjective assessment as the measure of student achievement. This is an ordinal variable, and for the ordinal dependent variable the ordered Probit or ordered Logit model should be the most suitable if there is no endogeneity problem. However, whether or not the Lewbel IV can be applicable to the ordinal variable is unclear yet. Thus, I transform the subjective assessment in two ways. (1) The subjective assessment have five categories, from ‘excellent’ to ‘very bad’, so I assign a score of 5 to ‘excellent’, 4 to ‘good’ and so on. Then just view it as a continuous variable. (2) Reclassify the students into two groups: excellent or not, then the linear probability model (LPM) can be used. In addition, I also tentatively try to combine the IV Probit model and Lewbel IV.

Columns (4) to (6) of Appendix Table C.1 present the summary statistics for the subjective assessment sample. The sample characteristics are nearly the same as the test score sample, although it has one more year’s observations. Table 4 presents the estimation results for the subjective assessment. These results have similar pattern with those for test score: Lewbel IV estimates are larger than the OLS/Probit estimates, but the effect is still modest — a 1000 yuan increase in the private tutoring expenditure only improve the subjective assessment by 0.09 point (less than 13 percent of the standard deviation), or increase the probability of being excellent by four (Lewbel IV + LPM)
Table 4: Robustness checks (2)

<table>
<thead>
<tr>
<th></th>
<th>Subjective assessment</th>
<th>Subjective assessment: being excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>Lewbel IV</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Private tutoring expenditure (1000 yuan)</td>
<td>0.008</td>
<td>0.087**</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>Kleibergen-Paap rk Wald F-statistic</td>
<td>—</td>
<td>40.81</td>
</tr>
<tr>
<td>Observations</td>
<td>1,780</td>
<td>1,780</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses, and are clustered at the person level. *** p<0.01, ** p<0.05, * p<0.10. All specifications control for per capita family annual income, parental education, parental working status, gender, number of siblings, fixed effects of school grade, school quality, residential city and city-grade average private tutoring expenditure. The average marginal effect (AME) is reported for the Probit and IV Probit models. Data source: The Rural-Urban Migration in China (RUMiC) survey.

or three (Lewbel IV + IV Probit) percentage points. These results confirm the findings when the schooling achievement is measured with test score.

8 Further discussions

8.1 Reverse causality and fully simultaneous equations model

Up to now, the endogeneity is assumed to be due to the omitted variables. However, there may also exist reverse causality — the lower school test scores a child gets, the more money the household may want to spend on hiring private tutoring for the child. If so, the model turns into a fully simultaneous equations system:

\[ s_i = \tau p(p^e_i) + \varphi_1'X_i + \varepsilon_{1i}, \varepsilon_{1i} = \theta_1V_i + U_{1i}; \]

\[ (p^e_i)^p = \tau^p s_i + \varphi_2'X_i + \varepsilon_{2i}, \quad \varepsilon_{2i} = \theta_2V_i + U_{2i}. \]

In this case, in addition to the identification assumptions (i)-(iii) mentioned in Section 4, the error term of the schooling achievement equation, (11), is also required to be heteroskedastic, namely, \( Cov(X, \varepsilon_1^2) \neq 0 \). Besides, there are two endogenous variables \((p^e_i)^p \) and \(s_i\) in two equations, so at least I need two variables that satisfy the third moment condition - \( Cov(Z, \varepsilon_1\varepsilon_2) = 0 \). Then, the Equation (10) implies the following
Table 5: Fully simultaneous equations system estimation results

<table>
<thead>
<tr>
<th>Eq. by Eq.</th>
<th>OLS</th>
<th>Lewbel GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score Eq.</td>
<td>Exp. Eq.</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
</tbody>
</table>

Panel 1: Chinese test score and private tutoring expenditure equations system

Private tutoring expenditure (1000 yuan) 0.115 0.089

| Mathematics test score | 0.009 | 0.026 |
|                        | (0.008) | (0.021) |

Koenker heteroskedasticity test statistic $^{a}$ — — 16.49** 19.84***

J-Statistic — — 2.422 [0.659]

Observations 1109 1109

Panel 2: Mathematics test score and private tutoring expenditure equations system

Private tutoring expenditure (1000 yuan) 0.194* 0.446*

| Mathematics test score | 0.013* | -0.011 |
|                        | (0.008) | (0.023) |

Koenker heteroskedasticity test statistic $^{a}$ — — 14.60** 16.36**

J-Statistic — — 7.835 [0.099]

Observations 1111 1111

Note: Huber-White robust standard errors are in parentheses. The numbers in square bracket behind J-statistic are the p-values. *** p < 0.01, ** p < 0.05, * p < 0.10. All specifications control for per capita family annual income, parental education, parental working status, gender, number of siblings, fixed effects of school grade, school quality, residential city and city-grade average private tutoring expenditure. Data source: The Rural-Urban Migration in China (RUMiC) survey.

$^{a}$ The Koenker heteroskedasticity test is basically the same as the Breusch-Pagan heteroskedasticity test but drops the assumption of normally distributed errors. Here the null hypothesis of this test is that the error terms of the test score or private tutoring expenditure equations is homoskedastic with respect to the city-grade average expenditure on private tutoring and the school grade dummies.

moment conditions:

$$E[X_i(s_i - \tau^p(p^pe_i^p) + \varphi'_1X_i)] = 0,$$

$$E[X_i(p^pe_i^p - \tau^s s_i + \varphi'_2X_i)] = 0,$$

$$E[Z_i - \mu] = 0,$$

$$E[(Z_i - \mu)(s_i - \tau^p(p^pe_i^p) + \varphi'_1X_i)(p^pe_i^p - \tau^s s_i + \varphi'_2X_i)] = 0,$$

where $\mu$ is the expectation of $Z$. Applying GMM to the above moments provides the estimate of $(\tau^p, \tau^s, \varphi'_1, \varphi'_2)$ (hereafter, Lewbel GMM).

As for the choice of $Z$, apart from the city-grade average private tutoring expenditure, I also use the school grades (corresponding 5 dummies). The school grade mainly depends on the birth cohort, and having controlling for the school grades, it is hard to imagine, e.g. students in grade 3 have systematically different innate abilities, motivations or something else contained in the unobserved common factors ($V_i$) from the a little younger or older
Table 6: The OLS, Lewbel IV and FE estimation results in the panel sample

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>OLS</th>
<th>Lewbel IV</th>
<th>FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Private tutoring expenditure (1000 yuan)</td>
<td>0.198</td>
<td>0.259</td>
<td>-0.231</td>
</tr>
<tr>
<td></td>
<td>(0.153)</td>
<td>(0.437)</td>
<td>(0.240)</td>
</tr>
<tr>
<td>Observations</td>
<td>578</td>
<td>578</td>
<td>578</td>
</tr>
<tr>
<td>Kleibergen-Paap rk Wald F-statistic</td>
<td>—</td>
<td>30.210</td>
<td>—</td>
</tr>
<tr>
<td>Mathematics</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Private tutoring expenditure (1000 yuan)</td>
<td>0.390***</td>
<td>0.801***</td>
<td>-0.083</td>
</tr>
<tr>
<td></td>
<td>(0.121)</td>
<td>(0.298)</td>
<td>(0.158)</td>
</tr>
<tr>
<td>Observations</td>
<td>574</td>
<td>574</td>
<td>574</td>
</tr>
<tr>
<td>Kleibergen-Paap rk Wald F-statistic</td>
<td>—</td>
<td>31.574</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: Robust standard errors are in parentheses, and are clustered at the person level. *** p<0.01, ** p<0.05, * p<0.10. All specifications control for per capita family annual income, parental education, parental working status, gender, number of siblings, fixed effects of school grade, school quality, residential city and city-grade average private tutoring expenditure. Data source: The Rural-Urban Migration in China (RUMiC) survey.

students. Thus, school grades should also satisfy the third moment condition.

Table 5 presents the fully simultaneous equations system model estimation results. Columns (1) and (2) are equation-by-equation OLS estimation results, and Columns (3) and (4) are Lewbel GMM estimation results. First, the significant level of heteroskedasticity is lower than that in the main empirical analysis. This is because the error terms in score and private tutoring expenditure equations are less heteroskedastic with respect to the city-grade average private tutoring expenditure and school grades, respectively. Second, Lewbel GMM estimation results indicate that private tutoring has no causal effect on Chinese test score and vice verse; for mathematics test score, the causal effect of private tutoring is lower than that in the main results, and only significant at 10% level, but more importantly there is no evidence of reverse causality, just as for Chinese test score. These results offer additional credibility to the findings in the main analysis.

8.2 Panel sample and fixed effects model estimation results

The RUMiC project is a longitudinal survey, thus it should be useful to get rid of the time-invariant unobserved factors. The reason that the main analysis does not make use of this is because the panel sample loses about half observations. However, it may still provide us some insights into the causal effect on student achievement. In this section, I first

---

15Lewbel GMM is implemented using the system GMM estimation in Eviews, which has only the Huber-White robust standard errors but cannot adjust the standard error by clustering.
replicate the OLS and Lewbel IV estimations in the panel sample, which are presented in Columns (1) and (2) of Table 6. The results are basically the same as in the full sample. The most notable difference may be that the association between mathematics test score and private tutoring expenditure estimated by OLS doubles and becomes more significant, and the causal effect estimated by Lewbel IV reduces by approximately one third.

Now, let us consider the fixed effects model (FE) estimation. The theoretical model shows that the unobserved common factors \( (V_i) \) lead to endogeneity problem. If the common factor and its effect on schooling achievement as a whole, \( \delta_1 V_i \), are time-invariant, then the FE model can provide a consistent estimate. Among these unobserved common factors, the innate ability should be stable over time, while the others might change over time. If \( \delta_1 V_i \) change only a little over time, the FE estimation can partially correct the omitted-variable bias. Furthermore, even if we don’t know what fraction of \( \delta_1 V_i \) is time-invariant, the FE estimation can still give us some insightful thoughts about the direction of bias of the OLS estimate.

Column (3) of Table 6 reports the FE estimate in the panel sample. Compared to the OLS estimates, the FE estimates become negative and statistically insignificant for both Chinese and mathematics test scores. The negative FE estimates may be because private tutoring has an adverse effect on student achievement. As mentioned in the section of theoretical model, the adverse effect is theoretically possible but it should not be a universal phenomenon, otherwise, one would expect that the number of private tutoring participants would become fewer and fewer. However, the reality is that private tutoring participation rate has increased to more than half in urban China during the past decade. Therefore, we can safely rule out this possibility.

Suppose that private tutoring has a positive effect on schooling achievement, then there are two other possible explanations for this discrepancy between OLS and FE estimates. First, measurement error is likely to be greater for the fixed effect model, so the attenuation bias will be greater for FE estimator and this causes the FE estimates be closer to zero. However, since the attenuation bias does not change the sign of the estimate, measurement error can only explain part of the difference between OLS and FE estimates.

Second, peer effects, teacher quality, parental effort and other omitted variables in the
achievement equation change over time. When the changes in these time-varying factors are negatively correlated with the change in private tutoring expenditure, and when this negative correlation is high enough, the FE estimate can become negative. Moreover, in this case, it is reasonable to believe that the correlation between unobserved time-varying factors and private tutoring expenditure is negative. This can be seen by the following informal derivation.

If \( \text{cov}(\Delta(p^p \epsilon_{it}^p), \Delta \epsilon_{1it}) < 0 \) is true, then \( \text{cov}(\Delta e_{it}^p, \Delta \epsilon_{1it}) < 0 \) is also true because \( p^p \) is positive and exogenous. Also,

\[
\text{cov}(\Delta e_{it}^p, \Delta \epsilon_{1it}) = \text{cov}(e_{it}^p - e_{i,t-1}^p, \epsilon_{1it} - \epsilon_{1i,t-1})
= \text{cov}(e_{it}^p, \epsilon_{1it}) - \text{cov}(e_{i,t-1}^p, \epsilon_{1it})
- \text{cov}(e_{it}^p, \epsilon_{1i,t-1}) + \text{cov}(e_{i,t-1}^p, \epsilon_{1i,t-1}).
\]

Among the four terms on the right-hand side, the first one and the last one should be about equal. The second and third terms may or may not be equal, but their magnitude is very likely to be lower than that of the other two terms. This is because households usually respond more to the current situation than past or future situations. In this case, \( \text{cov}(e_{it}^p, \epsilon_{1it}) < 0 \), namely the level of private tutoring expenditure is negatively correlated with the level of unobserved time-varying factors. For example, if a parent observes his/her child’s test score reduced in the last semester, he/she is more likely to buy more private tutoring service for the child in this semester.

If the unobserved time-varying factors are negatively correlated to private tutoring expenditure, it is likely so are the unobserved time-invariant factors, too. Thus, although the FE model does not provide a consistent estimate of the causal effect of private tutoring on schooling achievement, it does suggest that the private tutoring expenditure is more likely to be negatively related to the omitted variables, and hence the OLS estimates should be an underestimate of the causal effect. This is consistent with our finding from the comparison between the OLS and Lewbel IV estimates.

8.3 Instrumental variable quantile regression results

Private tutoring should have different effects for the ‘good’ and ‘bad’ students. In order to investigate this kind of heterogeneity, I also estimate the model using the quantile regression (QR) and instrumental variable quantile regression (IVQR) proposed by Cher-
Table 7: Quantile regression and instrumental variable quantile regression results

<table>
<thead>
<tr>
<th></th>
<th>Chinese QR</th>
<th>Chinese IVQR</th>
<th>Mathematics QR</th>
<th>Mathematics IVQR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>0.10th quantile</td>
<td>0.265</td>
<td>0.949***</td>
<td>0.118</td>
<td>1.904***</td>
</tr>
<tr>
<td></td>
<td>(0.287)</td>
<td>(0.322)</td>
<td>(0.282)</td>
<td>(0.362)</td>
</tr>
<tr>
<td>0.25th quantile</td>
<td>0.068</td>
<td>1.247***</td>
<td>0.019</td>
<td>1.092***</td>
</tr>
<tr>
<td></td>
<td>(0.161)</td>
<td>(0.241)</td>
<td>(0.193)</td>
<td>(0.254)</td>
</tr>
<tr>
<td>0.50th quantile</td>
<td>0.092</td>
<td>0.472**</td>
<td>0.264**</td>
<td>0.762***</td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.212)</td>
<td>(0.127)</td>
<td>(0.229)</td>
</tr>
<tr>
<td>0.75th quantile</td>
<td>-0.039</td>
<td>-0.253</td>
<td>0.214**</td>
<td>0.621**</td>
</tr>
<tr>
<td></td>
<td>(0.088)</td>
<td>(0.233)</td>
<td>(0.085)</td>
<td>(0.254)</td>
</tr>
<tr>
<td>0.90th quantile</td>
<td>-0.006</td>
<td>0.434</td>
<td>0.142**</td>
<td>0.544*</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.299)</td>
<td>(0.062)</td>
<td>(0.322)</td>
</tr>
</tbody>
</table>

Kleibergen-Paap rk Wald F-statistic

|                      | —          | 19.36        | —              | 20.16           |

Observations 1109 1109 1111 1111

Note: Standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10. All specifications control for per capita family annual income, parental education, parental working status, gender, number of siblings, fixed effects of school grade, school quality, residential city and city-grade average private tutoring expenditure. Data source: The Rural-Urban Migration in China (RUMiC) survey.

The Kleibergen-Paap rk Wald F-statistic is borrowed from the IV estimation in order to judge the strength of the instrumental variables.

The QR results indicate that the effect of private tutoring on test scores differs in different quantiles of the test score distribution. For Chinese test score, the coefficient is positive at the bottom end but negative at the top end, although it is neither statistically significant nor economically substantial in all five quantiles. For mathematics test score, the coefficient is positive in all five quantiles and significant in the 0.50th, 0.75th and 0.90th quantiles. However, the magnitude of the coefficient is not large in the all five quantiles, either.

The IVQR estimates of the causal effect of private tutoring on Chinese test score are statistically significant in the 0.10th, 0.25 and 0.50th quantiles and higher than the QR

---

16 The quantile regression is implemented with the Stata command ‘qreg’, while the instrumental variable quantile regression is implemented with the Stata program written by Do Won Kwak, ‘ivqreg’. The former does not provide a robust estimation for standard errors; the latter has a robust option but the algorithm to estimate the robust standard error often does not work. Thus only the unadjusted standard errors are presented here, and thus, the significance level might be overstated.

17 The KP F-statistic is borrowed from the IV estimation, as a particular weak IV test in the framework of IVQR has not been developed so far.
estimates in these three quantiles; for mathematics, the IVQR estimates are significant and larger the QR estimates in all the five quantiles, and the magnitude declines when moving upward to the top tail. These results suggest the effect of private tutoring on student achievement has strong heterogeneity over the test score distribution, and private tutoring is more likely to improve the test score at the bottom end of test score distribution. Specifically, a 1000 yuan increase in private tutoring expenditure (about 54% of a standard deviation) raises Chinese test score by 1.25 percentage points (about 18% of a standard deviation) in the 0.25th quantile, and enhance the mathematics test score by 1.90 percentage point (about 26% of a standard deviation) in the 0.10th quantile.

There may be two reasons for this heterogeneous impact over the test scores distribution. First, the effectiveness of learning may encounter an increasing marginal cost curve. Consequently, the marginal effect of private tutoring reduces with the increase of test score. Second but maybe more important, the students in the top end of test score distribution may be more likely to participate in the English or art/sport training, while the students in the bottom end of test score distribution may be more likely to attend private tutoring in Chinese and Mathematics. Due to the data limitation, I cannot directly test this, however, using the small survey I conducted in Shaanxi Province, I try to provide some indirect evidence. I estimate the relationships between private tutoring participations and expenditures in different subjects and student and family characteristics in the small survey using the seemingly unrelated regression, as shown in Table A.2.

The results show that if the father is more educated, the child is more likely to participate in art/sport training; mother’s education significantly decreases the private tutoring expenditures in Chinese and mathematics, but significantly increases the expenditure in art/sport training; and the family income has a larger effect on private tutoring in English and art/sport (especially in English) than in Chinese and mathematics. Given that these family characteristics are positively related to students test scores, as in the findings in my paper and previous studies, it is reasonable to believe that the students from families with better socio-economic status are more likely to have higher Chinese and mathematics test scores, and they are also more likely to participate in and spend more on English and art/sport training but less likely to participate in and spend less on Chinese and mathematics. As a result, the total private tutoring expenditure has greater
and more significant effects in the lower quantiles than in the upper quantiles.

9 Summary and conclusions

Private tutoring is increasingly popular all over the world. Understanding the effect of private tutoring on schooling achievement is an important research issue. Given that the students with a better socio-economic background are more likely to participate in private tutoring, this question is particularly interesting to policy makers. The equality of opportunity is one important object of modern society. If private tutoring can effectively improve schooling achievement, and therefore further improves final educational attainment, private tutoring might seriously obstruct equality of opportunity. If private tutoring has little effect on student achievement, a reduction in expenditure for private tutoring may enhance a household’s utility and even the whole society’s utility. However, the classical OLS estimation of the effect suffers from the potential endogeneity problem, and thus cannot provide a consistent estimate.

In this study, I first using a theoretical model illustrate that the endogeneity problem is due to omitted-variable bias and highlight the conditions under which the effect of private tutoring on schooling achievement can be identified. Then, I use a heteroskedasticity-based method proposed by Lewbel (2012) (Lewel IV) to handle the endogeneity problem. Compared to the conventional IV estimation, this novel method does not need excluded instrumental variables but relies on other two assumptions: (i) the error terms for private tutoring are heteroskedastic; (ii) there are some variables which are independent of the omitted variables or uncorrelated to the variance of the omitted variables.

Using this new identification and estimation method, I find that, on average, private tutoring expenditure has a modest and positive causal effect on mathematics test scores of primary school students. An 1000 yuan increase in private tutoring (i.e. 54% of a standard deviation) raises the mathematics test score of primary school students by 1.07 percentage point (i.e. 15% of a standard deviation). In contrast, private tutoring expenditure has no statistically significant effect on the Chinese test score of primary school students. These results are robust to whether or not the initial endowment (birth weight) is controlled for, whether or not formal education expenditure is endogenous and whether the test score or subjective assessment is used to measure student achievement.
I consider the possibility of endogeneity due to reverse causality, and the Lewbel GMM results confirm the findings in the main analysis. Using the panel sample, the fixed effects model help gauge the direction of the bias in the OLS estimation, which is downward biased, although it cannot eliminate the bias. Finally, by combining the Lewbel IV method with the instrumental variable quantile regression proposed by Chernozhukov and Hansen (2008), I also find that private tutoring mainly affect the students in the bottom end of the distributions of test scores. For mathematics, an additional 1000 yuan spent on private tutoring leads to about 1.9 percentage points increase in test score in the 0.10th quantile. For Chinese, the results are a little mixed, but they still suggest that private tutoring is more likely to improve the test score in the lower quantiles. This heterogenous impact over test score distribution may be because of an increasing marginal cost curve in the schooling achievement production, but it may also because the worse-performed students are more likely to attend private tutoring in Chinese and mathematics while the better-performed students are more likely to attend English and art/sport training.

Two questions arise from this study. Why are the effects of private tutoring on the mathematics and Chinese test scores different? And given the nil or modest effect of private tutoring, why has private tutoring become more and more popular? Due to the limitations of the data, I only provide some conjectures for these two questions: the content of private tutoring may not be very directly relevant to the final examination; children may be not willing to attend private tutoring, and therefore, not cooperative with tutors; the private tutoring expenditure is defined as the total expenditure in the data, and the actual effect may be larger than the estimates obtained here if we have precise information on private tutoring expenditure for different subjects. Thus, an interesting avenue for future research is to collect the relevant data and seriously test these conjectures.
Acknowledgement

I would like to express great appreciation to Professors Xin Meng, Dr. Tue Górgens and Dr. Paul Chen for their valuable and constructive suggestions during the planning and development of this research work. Their willingness to give their time so generously has been very much appreciated. I also appreciate the two anonymous referees for their insightful comments and suggestions, which are extremely helpful in improving the quality of this paper.

I would also like to thank Sen Xue and participants at the Econometric Society Australian Meeting in Melbourne, July 2012, and the seminars in Australian National University, Peking University, Xiamen University, Wuhan University, Shandong University, and Southwestern University of Finance and Economics. They offer me lots of helpful comments. I also owe heartfelt thanks to Ms. Yaning Zhao, Ms. Huiying Liang, MS. Yaling Gong and Ms. Shuaixia Zhu. They helped me a lot when I conducted the small survey to supplement the main data.

Finally, I gratefully acknowledge the support from the Start-Up Grant [2014-22] at Southwestern University of Finance and Economics.
References


A Data appendix

This appendix introduces the survey I conducted between October and December in 2012 in Shaanxi province.

The survey included three primary schools in Shaanxi province. Of these, one is in a prefecture-level city (called A, hereafter), one is in a county town (called B, hereafter) and the last one is a rural school. Because nearly no student in the rural school participated in private tutoring (the same as in the RUMiC data), I only present the private tutoring behaviors of the students in the two urban schools.

In the school A, one class from each of Grades 4, 5 and 6 was sampled; and in the school B, only one class in Grade 4 was sampled. For students in Grade 4, there are two questionnaires. One is for students themselves, which contains questions about gender and private tutoring participation, and students’ willingness to attend private tutoring; the other one is for parents, and contains the questions about basic demographic information, family background and private tutoring expenditure. For the students in Grades 5 and 6 in the school A, only the questionnaire survey for students was conducted. That is, for these students, we only know their gender, private tutoring participation and attitude to private tutoring. The summary statistics of the variables are presented in Table A.1, and the estimation results of the private tutoring participations and expenditures in different subjects using seemingly unrelated regression are presented in Table A.2.
Table A.1: Summary statistics for the private tutoring survey in Shaanxi Province

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Mean</th>
<th>S.D</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.46</td>
<td></td>
<td>197</td>
</tr>
<tr>
<td>Age</td>
<td>9.26</td>
<td>0.62</td>
<td>127</td>
</tr>
<tr>
<td>Number of siblings</td>
<td>0.36</td>
<td>0.59</td>
<td>127</td>
</tr>
<tr>
<td>Local urban hukou</td>
<td>0.77</td>
<td></td>
<td>127</td>
</tr>
<tr>
<td>Father’s years of schooling</td>
<td>12.73</td>
<td>2.67</td>
<td>127</td>
</tr>
<tr>
<td>Mother’s years of schooling</td>
<td>12.37</td>
<td>2.94</td>
<td>125</td>
</tr>
<tr>
<td>Monthly family income (yuan)</td>
<td>4033</td>
<td>3252</td>
<td>99</td>
</tr>
<tr>
<td>Grade 4</td>
<td>0.55</td>
<td></td>
<td>207</td>
</tr>
<tr>
<td>Grade 5</td>
<td>0.17</td>
<td></td>
<td>207</td>
</tr>
<tr>
<td>Grade 6</td>
<td>0.28</td>
<td></td>
<td>207</td>
</tr>
</tbody>
</table>

*Private tutoring participation*

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Any subject</td>
<td>0.79</td>
<td></td>
<td>196</td>
</tr>
<tr>
<td>Chinese</td>
<td>0.30</td>
<td></td>
<td>197</td>
</tr>
<tr>
<td>Mathematics</td>
<td>0.32</td>
<td></td>
<td>197</td>
</tr>
<tr>
<td>English</td>
<td>0.65</td>
<td></td>
<td>197</td>
</tr>
<tr>
<td>Art/sport</td>
<td>0.11</td>
<td></td>
<td>197</td>
</tr>
</tbody>
</table>

*Private tutoring expenditure (yuan)*

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>838</td>
<td>704</td>
<td>35</td>
</tr>
<tr>
<td>Mathematics</td>
<td>956</td>
<td>877</td>
<td>21</td>
</tr>
<tr>
<td>English</td>
<td>1292</td>
<td>850</td>
<td>75</td>
</tr>
<tr>
<td>Art/sport</td>
<td>1090</td>
<td>894</td>
<td>15</td>
</tr>
</tbody>
</table>

*Willingness to attend private tutoring*

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Very much</td>
<td>0.19</td>
<td></td>
<td>251</td>
</tr>
<tr>
<td>Willing</td>
<td>0.41</td>
<td></td>
<td>251</td>
</tr>
<tr>
<td>A little reluctant</td>
<td>0.30</td>
<td></td>
<td>251</td>
</tr>
<tr>
<td>Very reluctant</td>
<td>0.11</td>
<td></td>
<td>251</td>
</tr>
</tbody>
</table>

Note: The expenditures are measured in the price of year 2012.

*a* These summary statistics are only for the private tutoring participants, but not all of the participants report the private tutoring expenditure.

*b* This question is only asked to the private tutoring participants. The observation are at the person-subject level.
Table A.2: Private tutoring participations and expenditures in different subjects: Seemingly unrelated regression results

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Chinese (1)</th>
<th>Math (2)</th>
<th>English (3)</th>
<th>Art/sport (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Private tutoring participation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girl</td>
<td>-0.310***</td>
<td>-0.222**</td>
<td>-0.041</td>
<td>0.266***</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.093)</td>
<td>(0.101)</td>
<td>(0.096)</td>
</tr>
<tr>
<td>Grade 4 in School A&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.201**</td>
<td>-0.096</td>
<td>-0.118</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.093)</td>
<td>(0.101)</td>
<td>(0.097)</td>
</tr>
<tr>
<td>Grade 5 in School A</td>
<td>-0.260*</td>
<td>-0.056</td>
<td>0.252*</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
<td>(0.133)</td>
<td>(0.145)</td>
<td>(0.138)</td>
</tr>
<tr>
<td>Monthly family income (1000 yuan)</td>
<td>0.024*</td>
<td>0.017</td>
<td>0.021</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Father’s years of schooling</td>
<td>0.008</td>
<td>0.017</td>
<td>-0.019</td>
<td>0.046**</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.018)</td>
<td>(0.020)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Mother’s years of schooling</td>
<td>-0.015</td>
<td>-0.022</td>
<td>-0.017</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.016)</td>
<td>(0.017)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Observations</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>R-squard</td>
<td>0.220</td>
<td>0.158</td>
<td>0.138</td>
<td>0.239</td>
</tr>
<tr>
<td><strong>Private tutoring expenditure (yuan)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>-266.412**</td>
<td>-38.790</td>
<td>299.974</td>
<td>295.887*</td>
</tr>
<tr>
<td></td>
<td>(122.436)</td>
<td>(140.717)</td>
<td>(183.272)</td>
<td>(164.150)</td>
</tr>
<tr>
<td>Grade 4 in School A&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-78.910</td>
<td>-89.866</td>
<td>3.986</td>
<td>197.795</td>
</tr>
<tr>
<td></td>
<td>(118.412)</td>
<td>(136.093)</td>
<td>(177.249)</td>
<td>(158.756)</td>
</tr>
<tr>
<td>Grade 5 in School A</td>
<td>-168.415</td>
<td>76.865</td>
<td>956.636***</td>
<td>114.805</td>
</tr>
<tr>
<td></td>
<td>(176.972)</td>
<td>(203.397)</td>
<td>(264.906)</td>
<td>(237.268)</td>
</tr>
<tr>
<td>Monthly family income</td>
<td>33.110**</td>
<td>24.127</td>
<td>52.367**</td>
<td>36.203*</td>
</tr>
<tr>
<td>Father’s years of schooling</td>
<td>18.273</td>
<td>38.230</td>
<td>15.618</td>
<td>36.649</td>
</tr>
<tr>
<td></td>
<td>(23.298)</td>
<td>(26.776)</td>
<td>(34.874)</td>
<td>(31.236)</td>
</tr>
<tr>
<td>Mother’s years of schooling</td>
<td>-33.595*</td>
<td>-11.785*</td>
<td>4.005</td>
<td>54.350**</td>
</tr>
<tr>
<td></td>
<td>(20.036)</td>
<td>(23.027)</td>
<td>(29.991)</td>
<td>(26.862)</td>
</tr>
<tr>
<td>Observations</td>
<td>86</td>
<td>86</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>R-squard</td>
<td>0.185</td>
<td>0.097</td>
<td>0.254</td>
<td>0.244</td>
</tr>
</tbody>
</table>

Note: All regressions control for the parents’ working status, but nearly none of them is statistically significant in any equation. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

<sup>a</sup> The omitted group is the Grade 4 in School B. However, the Grade 6 in School A is also omitted due to multicollinearity in regressions.
B Appendix figure

Figure B.1: Test score vs. private tutoring expenditure

Note: The private tutoring expenditure is measured in year 2009 price. There are 1,135 observations including the outliers in the figure. The solid circles represent the observations identified as outliers by the block adaptively computationally efficient outlier nominators (BACON) algorithm proposed by Billor et al. (2000). The figure shows that, if an observation has either a test score lower than 45, or has a private tutoring cost higher than 13,000 yuan, the BACON algorithm identifies it as the outlier. Data source: The Rural-Urban Migration in China (RUMiC) survey.
## C Appendix tables

### Table C.1: Summary statistics

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Test score sample</th>
<th>Subjective assessment sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (1)</td>
<td>S.D (2)</td>
</tr>
<tr>
<td><strong>Schooling achievement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese test score</td>
<td>91.35</td>
<td>7.05</td>
</tr>
<tr>
<td>Mathematics test score</td>
<td>92.44</td>
<td>7.35</td>
</tr>
<tr>
<td>Subjective assessment&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.76</td>
<td>0.69</td>
</tr>
<tr>
<td>Subjective assessment: being excellent</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td><strong>Schooling expenditure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal education expenditure (yuan)</td>
<td>1729</td>
<td>3344</td>
</tr>
<tr>
<td>Private tutoring expenditure (yuan)</td>
<td>1085</td>
<td>1869</td>
</tr>
<tr>
<td>Private tutoring participation&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td><strong>The variable to construct Lewbel instrumental variable&lt;sup&gt;c&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City-grade average private tutoring expenditure (yuan)</td>
<td>1079</td>
<td>589</td>
</tr>
<tr>
<td><strong>Years composition of the sample</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 2008</td>
<td>0.58</td>
<td>1120</td>
</tr>
<tr>
<td>Year 2009</td>
<td>0.42</td>
<td>1120</td>
</tr>
<tr>
<td>Year 2010</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Individual and household characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child’s age&lt;sup&gt;b&lt;/sup&gt;</td>
<td>9.88</td>
<td>1.80</td>
</tr>
<tr>
<td>Father’s age&lt;sup&gt;b&lt;/sup&gt;</td>
<td>39.76</td>
<td>4.68</td>
</tr>
<tr>
<td>Mother’s age&lt;sup&gt;b&lt;/sup&gt;</td>
<td>36.99</td>
<td>4.10</td>
</tr>
<tr>
<td>Father’s years of schooling</td>
<td>12.33</td>
<td>2.87</td>
</tr>
<tr>
<td>Mother’s years of schooling</td>
<td>11.68</td>
<td>2.82</td>
</tr>
<tr>
<td>Family annual income per capita (yuan)</td>
<td>18858</td>
<td>12398</td>
</tr>
<tr>
<td>Father: unemployed, retired or housekeeper</td>
<td>0.07</td>
<td>1120</td>
</tr>
<tr>
<td>Father: paid employment</td>
<td>0.81</td>
<td>1120</td>
</tr>
<tr>
<td>Father: self-employment</td>
<td>0.12</td>
<td>1120</td>
</tr>
<tr>
<td>Mother: unemployed, retired or housekeeper</td>
<td>0.21</td>
<td>1120</td>
</tr>
<tr>
<td>Mother: paid employment</td>
<td>0.71</td>
<td>1120</td>
</tr>
<tr>
<td>Mother: self-employment</td>
<td>0.09</td>
<td>1120</td>
</tr>
<tr>
<td>Female</td>
<td>0.47</td>
<td>1120</td>
</tr>
<tr>
<td>Number of siblings</td>
<td>0.16</td>
<td>0.40</td>
</tr>
<tr>
<td>Birth weight (kg)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3.35</td>
<td>0.56</td>
</tr>
</tbody>
</table>

*Continued on the next page.*
<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Test score sample</th>
<th>Subjective assessment sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (1)</td>
<td>S.D (2)</td>
</tr>
<tr>
<td><strong>School quality and school grades</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School quality: the best</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>School quality: above average</td>
<td>0.55</td>
<td>0.54</td>
</tr>
<tr>
<td>School quality: average or below</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>Grade 1</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>Grade 2</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>Grade 3</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>Grade 4</td>
<td>0.18</td>
<td>0.17</td>
</tr>
<tr>
<td>Grade 5</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>Grade 6</td>
<td>0.18</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Note: The income and expenditure are both in the 2009 price. The main regression analysis uses the test score sample, while the subjective assessment sample is used for the robustness check. Data source: The Rural-Urban Migration in China (RUMiC) survey.

* The parental subjective assessment of children’s performance has five categories: excellent, good, average, bad and very bad. Here, I assign a score of 5 to ‘excellent’, 4 to ‘good’, and so on.

* These variables are not used in the main regression analysis. Ages and private tutoring participation rate describe the sample characterizations, while birth weight is used in the robustness check.

* In the Lewbel IV estimation, the city-grade average private tutoring expenditure is used to construct the instrumental variables, CIV, in the way proposed by Lewbel (2012).

* These categories are omitted as reference groups in regressions.
Table C.2: The first stage results of Lewbel IV estimations

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIV(^a)</td>
<td>0.400***</td>
<td>0.407***</td>
<td>0.461***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.091)</td>
<td>(0.091)</td>
<td>(0.072)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City-grade average private tutoring expenditure (1000 yuan)(^a)</td>
<td>-0.117</td>
<td>-0.017</td>
<td>-0.112</td>
<td>-0.009</td>
<td>-0.012</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>(0.137)</td>
<td>(0.136)</td>
<td>(0.136)</td>
<td>(0.107)</td>
<td>(0.102)</td>
<td></td>
</tr>
<tr>
<td>Formal education expenditure (1000 yuan)</td>
<td>0.019</td>
<td>0.032*</td>
<td>0.019</td>
<td>0.033*</td>
<td>0.059***</td>
<td>0.073***</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.016)</td>
<td>(0.017)</td>
<td>(0.019)</td>
<td>(0.018)</td>
<td></td>
</tr>
<tr>
<td>Per capita family annual income (1000 yuan)</td>
<td>0.013*</td>
<td>0.014**</td>
<td>0.013*</td>
<td>0.013**</td>
<td>0.008*</td>
<td>0.007*</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td></td>
</tr>
<tr>
<td>Father’s years of schooling</td>
<td>0.024</td>
<td>0.025</td>
<td>0.024</td>
<td>0.023</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.028)</td>
<td>(0.028)</td>
<td>(0.020)</td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>Mother’s years of schooling</td>
<td>0.075***</td>
<td>0.070***</td>
<td>0.074***</td>
<td>0.073***</td>
<td>0.061***</td>
<td>0.056***</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.028)</td>
<td>(0.028)</td>
<td>(0.021)</td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>Father: paid employment</td>
<td>0.451***</td>
<td>0.467***</td>
<td>0.483***</td>
<td>0.494***</td>
<td>0.255**</td>
<td>0.228*</td>
</tr>
<tr>
<td></td>
<td>(0.144)</td>
<td>(0.136)</td>
<td>(0.146)</td>
<td>(0.128)</td>
<td>(0.132)</td>
<td></td>
</tr>
<tr>
<td>Father: self-employment</td>
<td>0.511**</td>
<td>0.435**</td>
<td>0.557**</td>
<td>0.473**</td>
<td>0.378*</td>
<td>0.244</td>
</tr>
<tr>
<td></td>
<td>(0.237)</td>
<td>(0.216)</td>
<td>(0.218)</td>
<td>(0.201)</td>
<td>(0.188)</td>
<td></td>
</tr>
<tr>
<td>Mother: paid employment</td>
<td>-0.188</td>
<td>-0.244</td>
<td>-0.173</td>
<td>-0.224</td>
<td>-0.034</td>
<td>-0.072</td>
</tr>
<tr>
<td></td>
<td>(0.148)</td>
<td>(0.151)</td>
<td>(0.146)</td>
<td>(0.110)</td>
<td>(0.113)</td>
<td></td>
</tr>
<tr>
<td>Mother: self-employment</td>
<td>-0.031</td>
<td>0.012</td>
<td>-0.039</td>
<td>0.011</td>
<td>-0.005</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>(0.223)</td>
<td>(0.226)</td>
<td>(0.224)</td>
<td>(0.190)</td>
<td>(0.186)</td>
<td></td>
</tr>
<tr>
<td>Female student</td>
<td>0.290**</td>
<td>0.305***</td>
<td>0.298**</td>
<td>0.317***</td>
<td>0.185**</td>
<td>0.211**</td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.114)</td>
<td>(0.118)</td>
<td>(0.091)</td>
<td>(0.087)</td>
<td></td>
</tr>
<tr>
<td>Number of siblings</td>
<td>-0.346***</td>
<td>-0.299**</td>
<td>-0.343***</td>
<td>-0.290**</td>
<td>-0.360***</td>
<td>-0.318***</td>
</tr>
<tr>
<td></td>
<td>(0.123)</td>
<td>(0.122)</td>
<td>(0.123)</td>
<td>(0.098)</td>
<td>(0.094)</td>
<td></td>
</tr>
<tr>
<td>Year 2009</td>
<td>0.126</td>
<td>0.108</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.076)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 2010</td>
<td>0.197*</td>
<td>0.169</td>
<td>0.198*</td>
<td>0.165</td>
<td>0.235**</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td>-0.117</td>
<td>-0.123</td>
<td>(0.117)</td>
<td>(0.124)</td>
<td>(0.117)</td>
<td>(0.123)</td>
</tr>
<tr>
<td>School quality: above average</td>
<td>-0.108</td>
<td>-0.147</td>
<td>-0.111</td>
<td>-0.150</td>
<td>-0.175</td>
<td>-0.212</td>
</tr>
<tr>
<td></td>
<td>(0.192)</td>
<td>(0.196)</td>
<td>(0.191)</td>
<td>(0.194)</td>
<td>(0.150)</td>
<td>(0.151)</td>
</tr>
<tr>
<td>School quality: average or below</td>
<td>-0.145</td>
<td>-0.132</td>
<td>-0.146</td>
<td>-0.133</td>
<td>-0.206</td>
<td>-0.209</td>
</tr>
<tr>
<td></td>
<td>(0.204)</td>
<td>(0.207)</td>
<td>(0.203)</td>
<td>(0.206)</td>
<td>(0.158)</td>
<td>(0.161)</td>
</tr>
<tr>
<td>Grade 2</td>
<td>0.145</td>
<td>0.104</td>
<td>0.126</td>
<td>0.088</td>
<td>0.066</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>(0.190)</td>
<td>(0.204)</td>
<td>(0.189)</td>
<td>(0.203)</td>
<td>(0.131)</td>
<td>(0.141)</td>
</tr>
<tr>
<td>Grade 3</td>
<td>0.353*</td>
<td>0.313</td>
<td>0.344*</td>
<td>0.298</td>
<td>0.160</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>(0.194)</td>
<td>(0.206)</td>
<td>(0.192)</td>
<td>(0.204)</td>
<td>(0.132)</td>
<td>(0.137)</td>
</tr>
<tr>
<td>Grade 4</td>
<td>0.720***</td>
<td>0.571***</td>
<td>0.707***</td>
<td>0.556**</td>
<td>0.569***</td>
<td>0.422***</td>
</tr>
<tr>
<td></td>
<td>(0.221)</td>
<td>(0.221)</td>
<td>(0.220)</td>
<td>(0.219)</td>
<td>(0.156)</td>
<td>(0.152)</td>
</tr>
<tr>
<td>Grade 5</td>
<td>0.809***</td>
<td>0.836***</td>
<td>0.789***</td>
<td>0.812***</td>
<td>0.584***</td>
<td>0.599***</td>
</tr>
<tr>
<td></td>
<td>(0.244)</td>
<td>(0.259)</td>
<td>(0.241)</td>
<td>(0.257)</td>
<td>(0.164)</td>
<td>(0.169)</td>
</tr>
<tr>
<td>Grade 6</td>
<td>0.618***</td>
<td>0.540***</td>
<td>0.612***</td>
<td>0.532**</td>
<td>0.473***</td>
<td>0.411***</td>
</tr>
<tr>
<td></td>
<td>(0.225)</td>
<td>(0.230)</td>
<td>(0.223)</td>
<td>(0.228)</td>
<td>(0.155)</td>
<td>(0.155)</td>
</tr>
</tbody>
</table>

Continued on the next page.
Table C.2 (continued)

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Chinese&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Mathematics&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Subjective assessment&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>City fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.640 (0.541)</td>
<td>-0.647 (0.523)</td>
<td>-0.690 (0.537)</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>1.109</td>
<td>1.109</td>
<td>1.111</td>
</tr>
<tr>
<td>Cragg-Donald Wald F statistic</td>
<td>—</td>
<td>75.66</td>
<td>78.31</td>
</tr>
<tr>
<td>Kleibergen-Paap rk Wald F statistic</td>
<td>—</td>
<td>19.36</td>
<td>20.16</td>
</tr>
<tr>
<td>Koenker heteroskedasticity statistic&lt;sup&gt;b&lt;/sup&gt;</td>
<td>8.823***</td>
<td>—</td>
<td>16.08***</td>
</tr>
</tbody>
</table>

Note: Robust standard errors are in parentheses, and are clustered at the person level. *** p<0.01, ** p<0.05, * p<0.10. Data source: The Rural-Urban Migration in China (RUMiC) survey.

<sup>a</sup> The city-grade average private tutoring expenditure (Z) is the average of private tutoring expenditure of other students in the same city and in the same or adjacent grade. CIV is the constructed instrumental variable: \((Z - \bar{Z})\hat{\varepsilon}^2\), where \(\hat{\varepsilon}^2\) are the residuals in the regression for private tutoring expenditure controlling for Z.

<sup>b</sup> The Koenker heteroskedasticity test is basically the same as the Breusch-Pagan heteroskedasticity test but drops the assumption of normally distributed errors. Here the null hypothesis of this test is that the error term of private tutoring expenditure equation is homoskedastic with respect to the city-grade average private tutoring expenditure.

<sup>c</sup> The dependent variable in this table is private tutoring expenditure. The labels ‘Chinese’, ‘mathematics’ and ‘subjective assessment’ indicate the samples with valid Chinese and mathematics test scores or parental subjective assessment of children’s schooling performance.